



A method for estimating physical and economic food access at high spatial resolution

Florencio Campomanes V¹ · Michael Marshall¹ · Andrew Nelson¹

Received: 28 February 2023 / Accepted: 4 October 2023
© The Author(s) 2023

Abstract

Physical and economic access to food vary spatially. Methods to map that variability at high levels of spatial detail over large areas are scarce, even though suitable datasets and methods exist. Using open-access data for Ethiopia, we developed a method to map the disparities in physical and economic food access at 1-km resolution. We selected 25 access-related variables for 486 geo-located communities from the 2018 Ethiopian Living Standards Measurement Study to create a food access index (FAI). The index was based on a weighted summation of the 25 variables from a principal component analysis (PCA). We then extrapolated the FAI to the rest of Ethiopia using a generalized additive model (GAM) to produce a 1-km resolution FAI map and used that to describe the spatial variability of food access. Economic access had a heavier weight than physical access in the FAI reflecting the fact that proximity to food markets alone is insufficient if one cannot afford food. The GAM had an R^2 of 0.57 and a normalized root mean square error of 22.2% which are comparable to measures of model performance in studies that provided micro-level estimates of relative wealth. Peri-urban areas, representing 67% of the population, had relatively low food access, suggesting that these areas should be a priority for infrastructure or economic intervention. The scarcity of detailed spatial information on food access may limit the effectiveness of targeted policymaking to improve food security. The methodology developed in this study uses widely available and carefully selected datasets and can contribute to more spatially detailed estimates of food access in other countries.

Keywords Food prices · Food security · Sustainable development · Spatial analysis · Household surveys · Sub-Saharan Africa · Machine learning

1 Introduction

Food access – one of the dimensions of food security - refers to the physical and economic ability of people to acquire food. The physical aspect concerns transport options, ease of mobility, and the proximity and travel time to food-related services such as markets (Lê et al., 2015). Economic access is the ability to purchase food, usually driven by a combination of food prices, household income, and access to financial support (Lele et al., 2016). Food access also influences other dimensions of food security. For example, an increase in access to food may contribute to increased diet diversity as shown in China (Liu et al., 2014; Wang et al., 2017).

Food access, and the underlying economic and physical factors that influence food access, vary over space and over time (Bondemark, 2020). Detailed and accurate estimation of these multiple dimensions of food access variability is essential for food policy related decisions (Ploeg et al., 2015). We briefly review different methods to characterise and map food access, identify limitations, and propose a methodology that addresses them.

Food access has been defined as the time or distance to travel to retail supermarkets or grocery stores that offer healthy food options (Garcia et al., 2020; Jin & Lu, 2021; Kolak et al., 2018) or in general (Chen, 2017). This definition focuses on physical access and is calculated by distance-related measures such as network distance (Kolak et al., 2018) and by quantifying how many food suppliers are within a catchment area (Chen, 2017; Garcia et al., 2020; Jin & Lu, 2021). Strome et al. (2016) characterised physical access with three criteria (proximity, transportation access, and realized access) and stated that proximity (as defined

✉ Florencio Campomanes V
f.campomanes@utwente.nl

¹ Faculty of Geo-Information Science and Earth Observation, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

in Kolak et al., 2018) and transportation are insufficient to accurately estimate food access.

Most studies that mapped physical were conducted within cities, with very few at country (Deller et al., 2015) or multi-country (Pozzi & Robinson, 2008) level. Deller et al. (2015) constructed a county-level food access index in the United States based on the number of healthy food sources (farmer's markets, grocery stores, and supercentres) for every 1000 people per county while Pozzi and Robinson (2008) defined food access as travel time to markets or urban centres in 7 countries in Eastern Africa. However, solely relying on physical access does not fully capture food access as a food security component, because maximizing the use of the transport infrastructure may not be economically possible for all households, even if the physical infrastructure is good (Nelson et al., 2019). For example, living far away from supermarkets does not necessarily mean lower food access (Coveney & O'Dwyer, 2009).

Some studies investigated the relationship between physical access to food and socioeconomic factors. For instance, Garcia et al. (2020) found that a cluster of poor neighbourhoods were spatially correlated with lower food access. Similar results were obtained in a spatiotemporal analysis by Jin and Lu (2021) that showed a spike in the number of areas with low food access during the global economic recession of 2008, suggesting that deprived areas had lower food access. Food budget allocation within households was greatly affected by food prices. Households would thus rather travel farther for more affordable food (Lin et al., 2014).

Few studies have simultaneously considered both the physical and economic dimensions of access to food. Breyer and Voss-Andreae (2013) and Jiao et al. (2012) found that food price and proximity to food sources were essential for measuring access to food but also acknowledged that this relationship required further investigation in other contexts. Koh et al. (2019) built a spatially explicit agent-based model that considered several factors (monthly shopping frequency, ability to carry items from the market to home, and probability of shopping in a general supermarket versus a convenience store/partial market) to investigate the impacts of several policy interventions to food access. Jiao et al. (2012) emphasized the importance of mobility, specifically, vehicle ownership and driving distances in the context of both physical and the economic access to food. Furthermore, Sharkey and Horel (2008) stated that for rural areas, distance to food retail can be a proxy for the cost of driving while in urban areas with low car ownership rates and high poverty, access to public transit and walking infrastructure are more important measures. Hemerijckx et al. (2022) studied the urban food accessibility of Kampala, Uganda, from both physical and economic standpoints separately. They equally weighted travel time to food sources, formal food system potential, and percentage of households involved in agricultural activities

to define physical access while economic access was defined as both the daily food expenditure per person and percentage of income for non-food items, also equally weighted. Hemerijckx et al. (2022) thus found that income is the primary food access constraint in most areas, while living far away from urban centres was less important. These studies, however, were conducted over a limited spatial domain (one town or city) or at coarse spatial resolution. Conducting studies across the urban–rural continuum is important too because both the physical and the economic access to food may vary substantially across urban, peri-urban, and rural areas (Janda et al., 2021; Nyangasa et al., 2019).

Measuring and mapping the multidimensional nature of food access is challenging. Previous studies have modelled either the food consumption score (FCS) or the reduced coping strategies index (rCSI) for one or more countries (Deléglise et al., 2022; Lentz et al., 2019; Martini et al., 2022). FCS and rCSI are common food insecurity severity measures but do not provide insight into the determinants of food insecurity. Further, these studies map results at cluster locations so they are not complete assessments, or at sub-national administrative units, which miss potential important spatial variations in food access that do not align with administrative boundaries. More spatially detailed representations of food access are required to better target policies, especially considering urban vs peri-urban vs rural settings, requiring an alternative approach to mapping food access.

Chi et al. (2022) quantified and mapped relative wealth with an asset-based wealth index. The index was constructed from survey data and represents the variation in wealth from the national average. The index was then extrapolated from survey locations with satellite image and other geospatial data to estimate wealth at 2.4 km resolution for 135 low- and middle- income countries. The availability of similar survey data and spatial datasets pertaining to food access suggest that a similar approach is feasible to create national-level, spatially detailed maps of economic and physical food access.

Our main aims were to develop a food access index based on commonly available household survey data, to create (micro) estimates of food access at high spatial resolution by extrapolating the food access index using commonly available geospatial data, and to demonstrate a use case by examining disparities in food access across the rural–urban spectrum. We use Ethiopia as a case study country because its food systems are spatially heterogenous (Esayas et al., 2018), vulnerable to climate shocks and food price fluctuations, (Belachew et al., 2012), and face frequent food insecurity (Mohamed, 2017). In the case of Ethiopia, we identify which household survey data variables are most highly correlated to food access, which geospatial variables most strongly contribute to quantifying food access and how these relate to food access, and how food access varies across along the urban–rural spectrum.

2 Material and methods

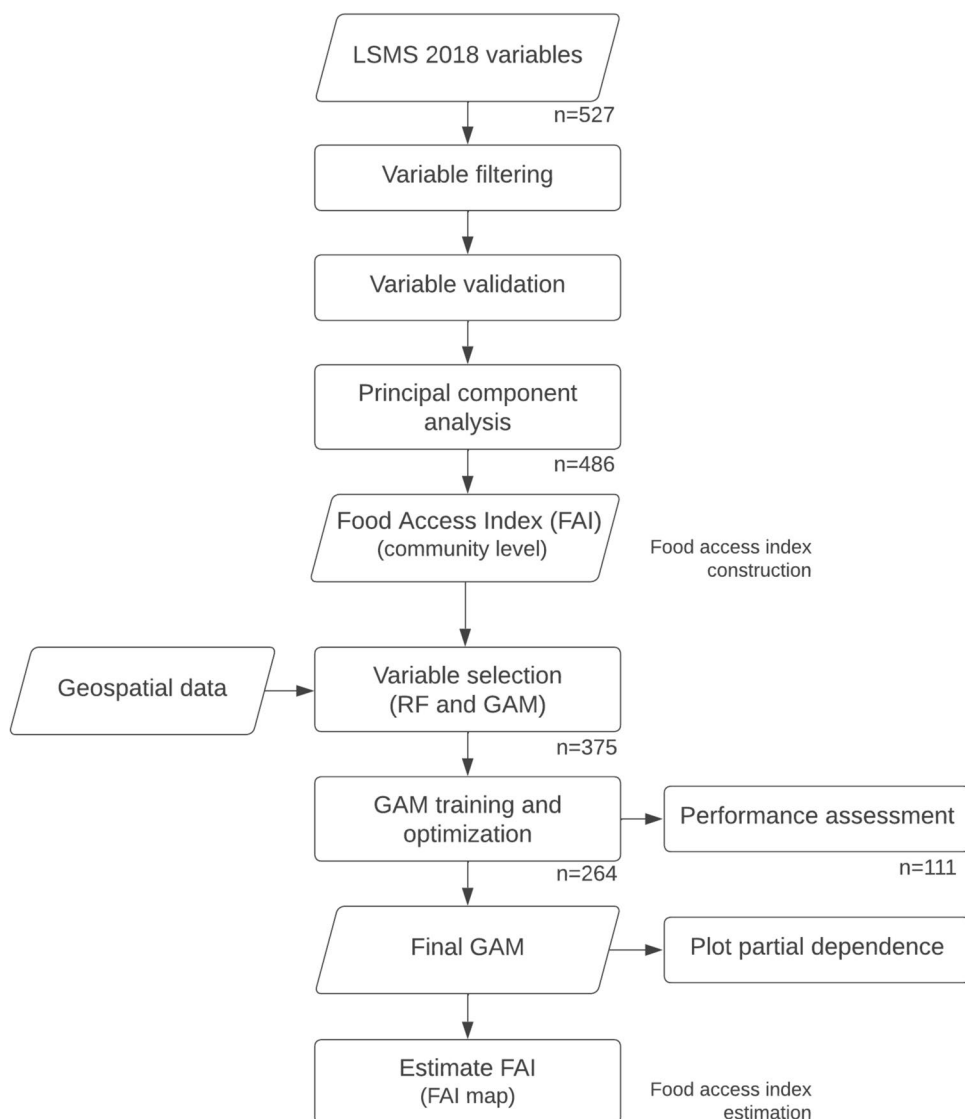
An overview of the methodology to quantify and map food access is shown in Fig. 1. There are two main steps in producing a nationally complete, high spatial resolution map of food access. The first identifies, filters, and then weights variables from available survey data to create a community-level food access index (FAI) for geolocated communities across the country. The second extrapolates the FAI to all other populated areas of the country using geospatial datasets related to food access to create the high resolution, national level map. A third and final step demonstrates the use of the map for describing spatial disparities in food access.

2.1 Estimating food access at community level

2.1.1 Identifying food access-related variables from household surveys

The 2018 Living Standards Measurement Survey (LSMS) in Ethiopia (Central Statistics Agency of Ethiopia, 2020) was conducted by the World Bank and the LSMS team between September 2018 and August 2019. The survey covers 6,804 households in 535 geolocated communities that were nationally representative of urban and rural areas (Central Statistics Agency of Ethiopia & World Bank, 2020). We identified 28 variables in the LSMS that were related to physical and economic food access. These included access to basic

Fig. 1 Methodology overview



services, asset ownership, access to credit, market prices, and topography. Ownership of assets has been found to have a positive relationship with wealth (Rutstein & Johnson, 2004) and in effect, increases food purchasing power. Other specific assets were included based on their relevance to food access. Among these, car ownership has been deemed relevant (Jiao et al., 2012) but there are several transport options in Ethiopia, so we included all transport-related assets. Also, access to electricity has been found to have a positive impact on food security (Candelise et al., 2021) and therefore we included ownership of electronic assets such as refrigerators. Mobile phone use and connectivity, in general, have been shown to have an important role in food access (Wantchekon & Riaz, 2019) hence the inclusion of mobile phone ownership. Access to formal credit has been found to improve food security (Salima et al., 2023) so we included variables related to access to financial assistance programs. Lastly, food prices were also considered as important variables (Lin et al., 2014).

The 28 variables required processing before further analysis. Variables for asset ownership that were collected at either individual or household level were aggregated to community level by calculating the percentage of ownership of each asset in each community. We removed redundant variables based on multicollinearity. We used cross-correlation analysis to identify pairs of highly correlated variables (pairs of variables with a Pearson's correlation coefficient > 0.8 , (Mason & Perreault, 1991)) and removed the highly correlated variable which had the highest Variable Inflation Factor (VIF, a measure of how much the variance of an explanatory variable is inflated by its correlation with other explanatory variables (Forthofer et al., 2007; Kim, 2019)). This process was repeated until there were no pairs of variables with a Pearson's correlation coefficient > 0.8 , resulting in 25 variables (Table 1). A total of 486 (out of 527) communities that had complete data for these 25 variables were retained for the next steps.

2.1.2 Validating the selected variables against independent measures of food access

Although these variables have been shown to relate to food access, none of them directly measure food access or measure the degree of food access that an individual, household or community has. Their inclusion in our model requires validation with independent data on food access, even though such data is limited. We hypothesise that there would be significant differences in the mean values of our selected variables (at household level) between households that experienced food insecurity due to food access and those that did not.

The Food Insecurity Experience Scale (FIES) Survey Module of LSMS includes households that reported food

insecurity in the seven days prior to the survey. Of the 6770 households in the LSMS dataset, 885 responded yes to experiencing food shortage in the FIES module. We extracted four variables from FIES and created an additional one (variables 26–30 in Table 1). One FIES variable reports whether a household had insufficient food. The other three variables concerned access-related causes of food insecurity specifically if the food in the market was too expensive, if the fare to the market was too expensive, and if the market was too far. An additional variable was calculated which represented whether food shortage was experienced because of any of the three access-related causes. Some 314 households reported this. We used a Mann–Whitney U test (Mann & Whitney, 1947) to determine if there was a significant difference in the means.

2.1.3 Constructing the food access index at community level

The food access index (FAI) was based on a weighted summation of the 25 variables where the variable weights were determined using principal component analysis, (PCA) (Hotelling, 1933) a methodology first proposed by Filmer and Pritchett (2001) and applied in similar approaches such as the Demographic and Health Survey (DHS) relative wealth index (Rutstein & Johnson, 2004). PCA is a widely used multivariate dimensionality reduction technique that produces new uncorrelated variables through orthogonal transformations that consist of linear combinations of the original input variables. A covariance matrix of the variable set is calculated from which a set of eigenvectors and eigenvalues are computed. Eigenvectors represent the coefficients or weights of each variable to a specific component while eigenvalues are the amount of variation in the data is accounted for. The linear combination that accounts for the maximum amount of variation is the first principal component (PC).

The 25 variables were mean centred and standardized since PCA is sensitive to the range of values in the data. An orthogonal Varimax rotation was used to ensure that each PC had only a few high loading variables, which aids in the interpretability of each PC. Only variables with significant coefficients were kept in the analysis to account for the sensitivity of the model to noise. Hair et al. (1998) stated that a coefficient with an absolute value greater than or equal to 0.3 is significant for a sample size of greater than 350. The standardized values of the significant variables were multiplied by their respective coefficients from the first PC and then summed per community following the methodology used to construct the DHS-based relative wealth index (Rutstein & Johnson, 2004). The weighted sums were then converted to z-scores to represent FAI per community.

Table 1 Ethiopia 2018 LSMS variables used in study

#	Category	Collection Level	Variable	Variable description	References
1	Characteristics	Community	slope	Slope of majority of the land	Yamaguchi et al. (2019)
2	Access to basic services	Community	road_type	Type of main access road (0–3, worst to best)	Wudad et al. (2021)
3			dist_bus	Distance to (in meters): • nearest bus station	Bergau et al. (2022)
4			dist_markets	• large weekly market Fare to (in Ethiopian Birr):	
5			fare_MUC	• nearest major urban centre (regional or zonal capital)	Bergau et al. (2022)
6			fare_woreda	• nearest woreda town	
7			Asset ownership	Household	
8	bike_perc	• Bicycle			Candilise et al. (2021)
9	motor_perc	• Motorcycle			Chen and Clark (2016)
10	cart_h_perc	• Cart – hand-pushed			Rutstein and Johnson (2004)
11	cart_a_perc	• Cart – animal-drawn			
12	car_perc	• Private car			
13	ker_perc	• Kerosene stove			
14	cyl_perc	• Cylinder stove			
15	ene_perc	• Energy-saving stove			
16	refr_perc	• Refrigerator			
17	shelf_perc	• Shelf for food storage			
18	bio_perc	• Biogas stove			
19	Access to assistance	Individual			phone_perc
19		Household	total_assist	Total value of: • Cash received • Food received • In-kind assistance received	Salima et al. (2023)
20	Market prices	Community		Price of crop group in nearest market	
21			price_1	• Cereals	Lin et al. (2014)
22			price_2	• Pulses	
23			price_3	• Oil seed	
24			price_4	• Vegetables	
25			price_5	• Fruits	
25	price_6	• Tubers			
26	Food Insecurity Experience Scale (FIES)	Household	FI	• Faced a situation with insufficient food	
27			FI _{expense}	• FI cause: food in market was too expensive	
28			FI _{transpoCost}	• FI cause: fare to market was too expensive	
29			FI _{distance}	• FI cause: market too far from village	
30			FI _{access}	• FI caused by lack of access	

2.2 Extrapolating the FAI to all populated areas

2.2.1 Geospatial datasets related to food access

The second step in our method relates the community level FAI with continuous geospatial variables at the same 486 community locations and uses that relationship to extrapolate the FAI to all other populated locations in the country.

We considered geospatial datasets based on their relevance to food access leading to 45 relevant variables (Table 2).

Road proximity relates to access to the land transport network and physical access to food. We computed the Euclidean distance from each community to the nearest road from OpenStreetMap data and aggregated the distance in to 12 ranges (Table 3). The cost of food is related to economic access. We computed an annual average cereal price by

Table 2 Geospatial data related to food access for extrapolating the FAI

Dataset	Source	# of variables	Relation to food access (and food security in general)
Road proximity	OpenStreetMap (OpenStreetMap contributors, 2022)	1	Physical access to road network and markets, and cost of travel
Cereal prices	Food price data from (World Food Programme, 2021)	1	Cost of food and economic access
Crop production	MapSPAM (International Food Policy Research Institute, 2020)	22	Locations of food production
Land use / land cover	ESA (Zanaga et al., 2021)	1	Relates to locations of food production and human settlements
Elevation	FABDEM (Hawker et al., 2022)	1	Constraint on physical access, also relates to food production
Bioclimatic variables	WorldClim 2 (Fick & Hijmans, 2017)	19	Climatic factors relate to food production
Population	WorldPop (WorldPop et al., 2018)	1	

modifying the methodology of Cedrez et al. (2020). While they used food price data from several sources and covered several countries, their data included food prices from only four markets in Ethiopia. Our modification used crop market prices from the World Food Programme (World Food Programme, 2021) which contained monthly food price data for several crops from 2000 to 2021 for 110 markets across Ethiopia. We used this to calculate an average cereal retail price in each market. This price was adjusted for inflation by dividing it with the consumer price index (CPI) of that year. We then trained a random forest (RF) (Breiman, 2001) model to estimate the CPI-adjusted cereal price for Ethiopia from population, precipitation, and market access. RF is a machine learning algorithm that leverages weak-learning regression trees to improve performance and has been used for such variable selection tasks since it is capable of computing variable importance. The remaining geospatial variables on crop production, land use, elevation, and climate mainly relate to the location of food production (food availability) an indirectly to the physical access and economic cost associated with connecting locations of food production to location of food consumption. All variables were clipped to the national boundary, resampled to 1-km spatial

resolution, and projected to a common WGS 84 Mercator projection. After which, pixels of uninhabited areas (with population of zero based on WorldPop et al., 2018) were omitted from all the datasets.

2.2.2 Geospatial data preparation

The geospatial datasets also required pre-processing before they could be spatially joined with the community-level FAI. Random spatial offsets are applied to all LSMS community geolocation data (up to 5-km for urban and 10-km for rural communities) for privacy and ethical reasons (Central Statistics Agency of Ethiopia & World Bank, 2021) following the approach by Burgert et al. (2013). To address this, we adapted the methodology of Chi et al. (2022) by calculating the population-weighted average of each geospatial variable in spatial buffers of 5×5 km grids around urban communities and 10×10 km grids around rural communities. Since the geospatial datasets were masked for populated areas, some of these 5×5 or 10×10 grids had unpopulated areas. To avoid bias in the population weighted average, we discarded 111 communities (375 were retained) where less than half of their 1 km pixels (< 13 for 5×5 and < 50 for 10×10) were unpopulated. We then removed redundant geospatial variables based on multicollinearity following the same procedure in Sect. 2.2.1. 25 geospatial variables were retained and were spatially joined to 375 communities. This dataset was split into 70% (264 communities) for variable selection and training and 30% (111) for testing the extrapolation model. We ensured that this split was the same for communities labelled as urban or rural in the LSMS.

Only a subset of predictors variables was used for scaling out to prevent over-fitting. The predictor variables were evaluated with the R package “Variable selection using random forest” v.1.1.0 (VSURF) (Genuer et al., 2010). VSURF uses

Table 3 Road proximity classes

Class	Distance range (meters)	Class	Distance range (meters)
1	0 – 100	7	1500 – 2000
2	100 – 250	8	2000 – 2500
3	250 – 500	9	2500 – 3000
4	500 – 750	10	3000 – 5000
5	750 – 1000	11	5000 – 7500
6	1000 – 1500	12	> 7500

several iterations of RF in three phases (thresholding, interpretation, prediction) for variable reduction. Each phase has a different emphasis: the thresholding phase (50 iterations) assess the variance of variable importance and removes irrelevant (low varying) variables, the interpretation phase (25 iterations) identifies the threshold at which incorporating additional predictors provides little to no explanatory power, and the prediction phase (25 iterations) optimizes the model in a stepwise fashion to minimize model error.

2.2.3 FAI extrapolation

A generalized additive model (GAM) (Hastie & Tibshirani, 1986) in the R package “mgcv” v.1.8-35 (Wood, 2011) was used to extrapolate the FAI to all populated areas of Ethiopia at 1 km resolution. This step includes training, parameter optimisation, and testing. Variables in the training step were removed if they were insignificant ($p > 0.01$). The smoothing parameters of each input variable were optimized by testing three smoothing functions for each variable: cubic regression splines (Wood, 2017), thin-plate regression splines (Wood, 2003), and P-splines (Eilers & Marx, 1996). The best model configuration was the one which produced the GAM with the lowest Akaike Information Criterion (AIC) and where all variable-smoothing function pairs were significant ($p < 0.01$). The best performing GAM was evaluated against the test dataset set using the coefficient of determination (R^2) and normalized root mean square error (nRMSE) to show the error with respect to the prediction range of the GAM. Partial dependence plots (PDPs: Friedman, 2001) were used to visualize the relationship between the FAI and each variable.

2.3 Measuring disparities in food access across the rural–urban spectrum

We demonstrate the use of the 1 km map of FAI by using it to characterise the variability in food access across urban, peri-urban, and rural areas considering a hierarchy of urban settlement sizes. Two studies have previously demonstrated that such information can inform food security policy and our aim here is to show that FAI maps can provide this information nationally with sufficient detail for spatially targeted policy interventions. Dean and Sharkey (2011) modelled fruit and vegetable intake in Texas, USA, and found that distance to supermarkets was important in urban areas but not in rural areas. They recommended interventions should consider the distance to retail food environments as an important factor for improving fruit and vegetable intake in rural areas. Nguyen et al. (2021) quantified processed-food intake at household level across the rural–urban spectrum in Viet Nam and found peri-urban and rural households had consumed more ultra-processed food compared to urban households.

We calculated the average FAI and total population within each Urban Rural Catchment Area (URCA) category. URCA represents the connection between these urban centres and their surrounding rural area. They are defined in Cattaneo et al. (2021) as the catchment areas of urban centres of different sizes and reveal whether a location (and its associated population) is 1, 2, 3 or more hours distant from a large, intermediate, or small urban centre. If there are large differences in FAI between URCA categories, then this suggests that these locations could benefit from interventions to improve food access.

Table 4 Group means and significant differences between means

<i>Household</i>			<i>Community</i>		
<i>Variable</i>	<i>Mean (FI_{access})</i> (n = 359)	<i>Mean (FS)</i> (n = 5872)	<i>Variable</i>	<i>Mean (FI_{access})</i> (n = 104)	<i>Mean (FS)</i> (n = 203)
*Total assistance (ETB)	1060	342	*Distance to nearest bus station	2.87	6.62
*Phone	0.7160	1.1800	*Fare to nearest woreda town	7.08	15.40
*Refrigerator	0.0557	0.1910	Fare to nearest major urban centre	29.16	39.81
*Shelf for storage	0.1590	0.2530	Road type (0–3, worst to best)	1.80	1.56
*Energy-saving stove	0.0696	0.1240	Oil seed price	32.82	33.30
*Car	0	0.0223	Cereal price	18.17	18.46
*Bike	0	0.0204	Fruit price	27.19	24.30
*Cylinder stove	0	0.0153	Slope	3.01	2.91
Kerosene stove	0.0446	0.0700	Distance to nearest weekly market	4.88	3.99
Cart (hand-pushed)	0.0084	0.0163	Pulse price	31.39	33.10
Biogas stove	0	0.0009	Vegetable price	22.95	23.30
Cart (animal-drawn)	0.0167	0.0194	Tuber price	13.55	13.45
Motorcycle	0.0084	0.0097			

* $p < 0.05$

3 Results

3.1 Validating the variables against independent measures of access-related food insecurity

We categorised households and communities into three types: those that (1) experienced food insecurity due to lack of access (FI_{access}), (2) experienced food insecurity due to non-access causes (FI), and (3) were food secure (FS).

Out of 13 household-level variables, 8 had significant differences in their means ($p < 0.05$) between FI_{access} households and FS households (Table 4). The means for all asset-ownership variables for FS households were greater than those of FI_{access} households while it was the opposite for total assistance received. Only 2 of the 12 community variables had significant ($p < 0.05$) differences between FI_{access} and FS communities. Except for distance to the nearest weekly market, the means of distance and fare variables were consistently lower for FI_{access} communities. The mean road type, quantified from 0–3 representing worst to best, was higher for FI_{access} communities. In terms of market prices, there was no clear trend in mean differences between FI_{access} and FS communities.

3.2 Community-level FAI

The first PC accounted for 18.3% of the variance while the second accounted for 6.9%. 14 variables were significant in the first PC (Fig. 2), which was a combination of variables

related to asset ownership, quality of roads, and fruit price (strong positive weights), and to distance and fare to services (weak negative weights). Of the 14 significant variables, 7 had absolute value coefficients greater than 0.5. These were related to asset ownership (5), road quality (1), and fare to services (1).

The FAI was calculated as the weighted sum of the 14 significant variables from the PCA. FAI is a z-score; a FAI of 2 means that a community has a level of food access two standard deviations higher than the national average. The FAI ranged from -2.9 to 3.2 across the 486 communities (Fig. 3). The larger positive range is due to more variables in the PCA with positive and stronger coefficients than those with negative coefficients. A large proportion of the communities had negative FAI values, mostly between -1 and 0. Urban centres had positive FAI values (Fig. 4).

3.3 Extrapolating the FAI to all populated places

The 45 potential geospatial variables were reduced to three in the final GAM. Firstly we removed 20 due to high multicollinearity (13 were bioclimatic variables, 6 were related to crop production, and elevation). VSURF then selected 5 variables for the final prediction phase based on whether their inclusion in a RF model would reduce prediction errors; road proximity (RPX), cereal price (CRP), precipitation in coldest quarter (B19), isothermality (B03), and barley production (BRL). Finally, barley production and isothermality were

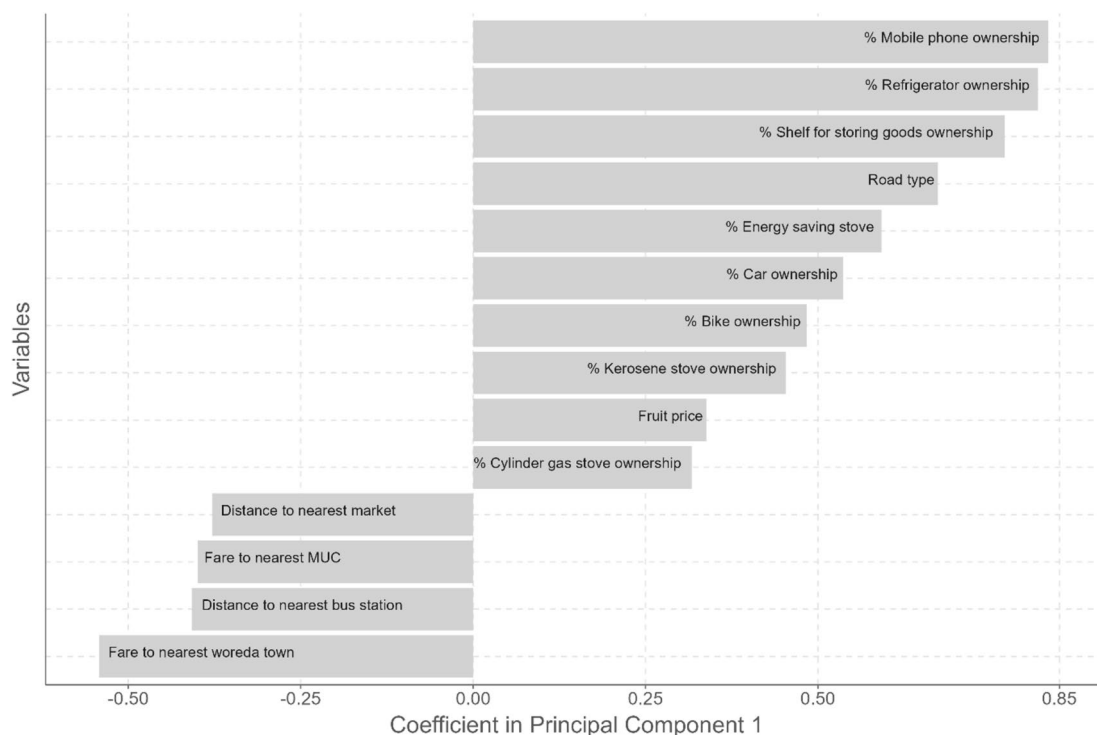


Fig. 2 Bar plot of the coefficients (loading) of the significant variables of the first PC

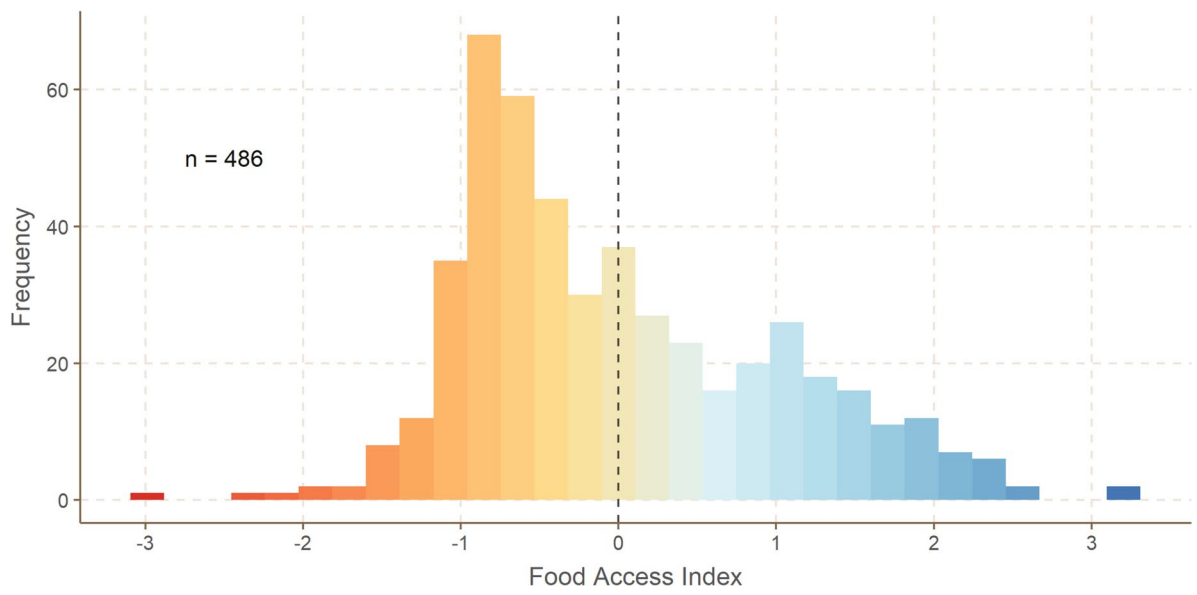


Fig. 3 Histogram of the community-level FAI

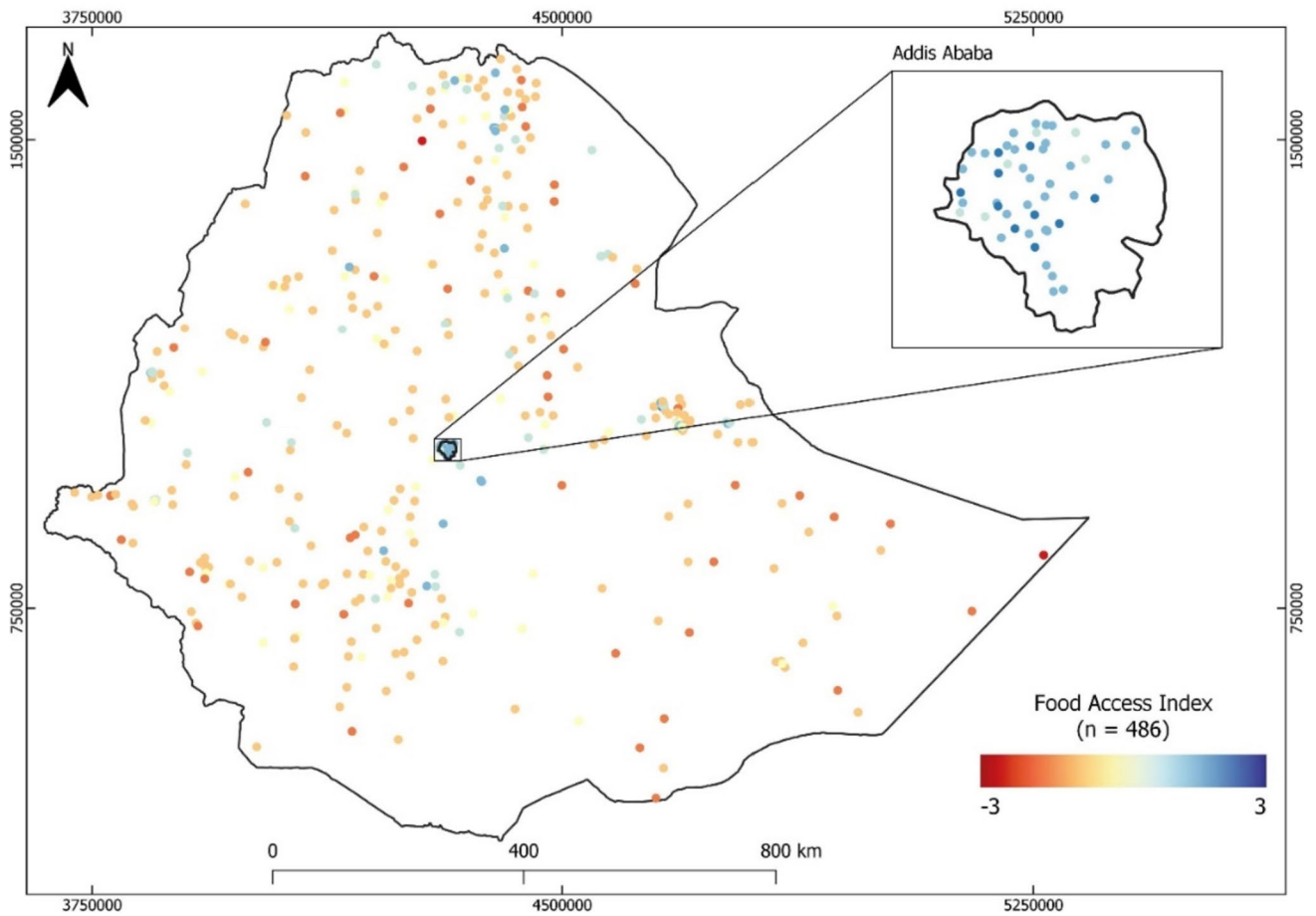


Fig. 4 Map of the community-level FAI

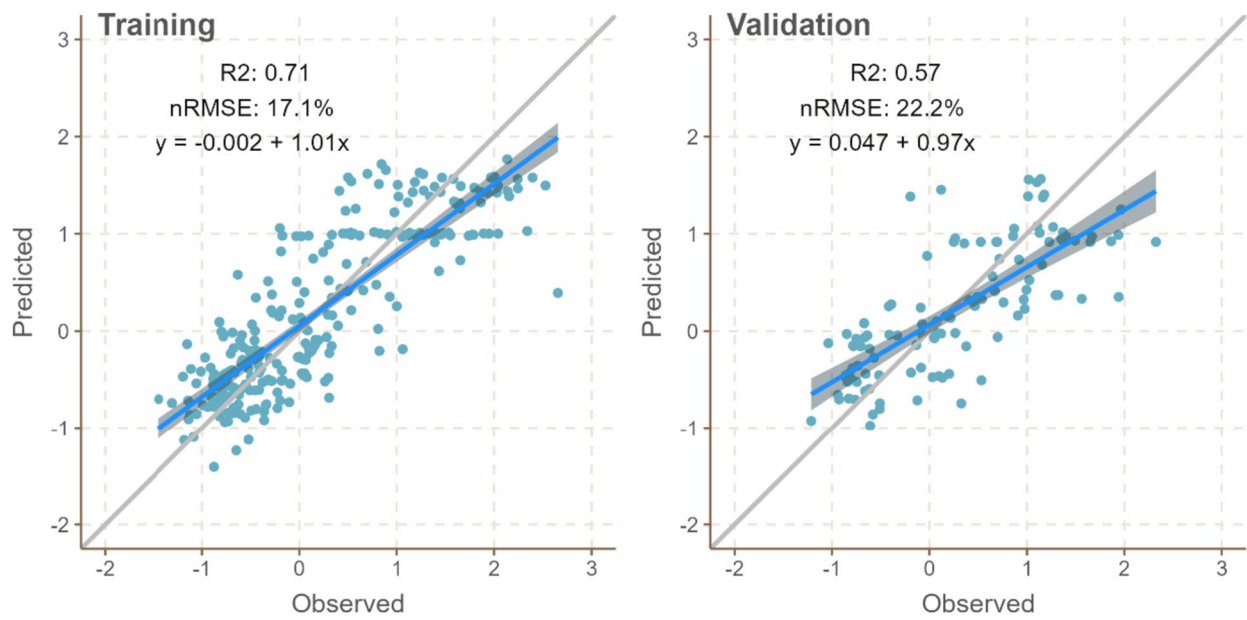


Fig. 5 Predicted-observed plot of best performing GAM on the training (left) and validation (right) set

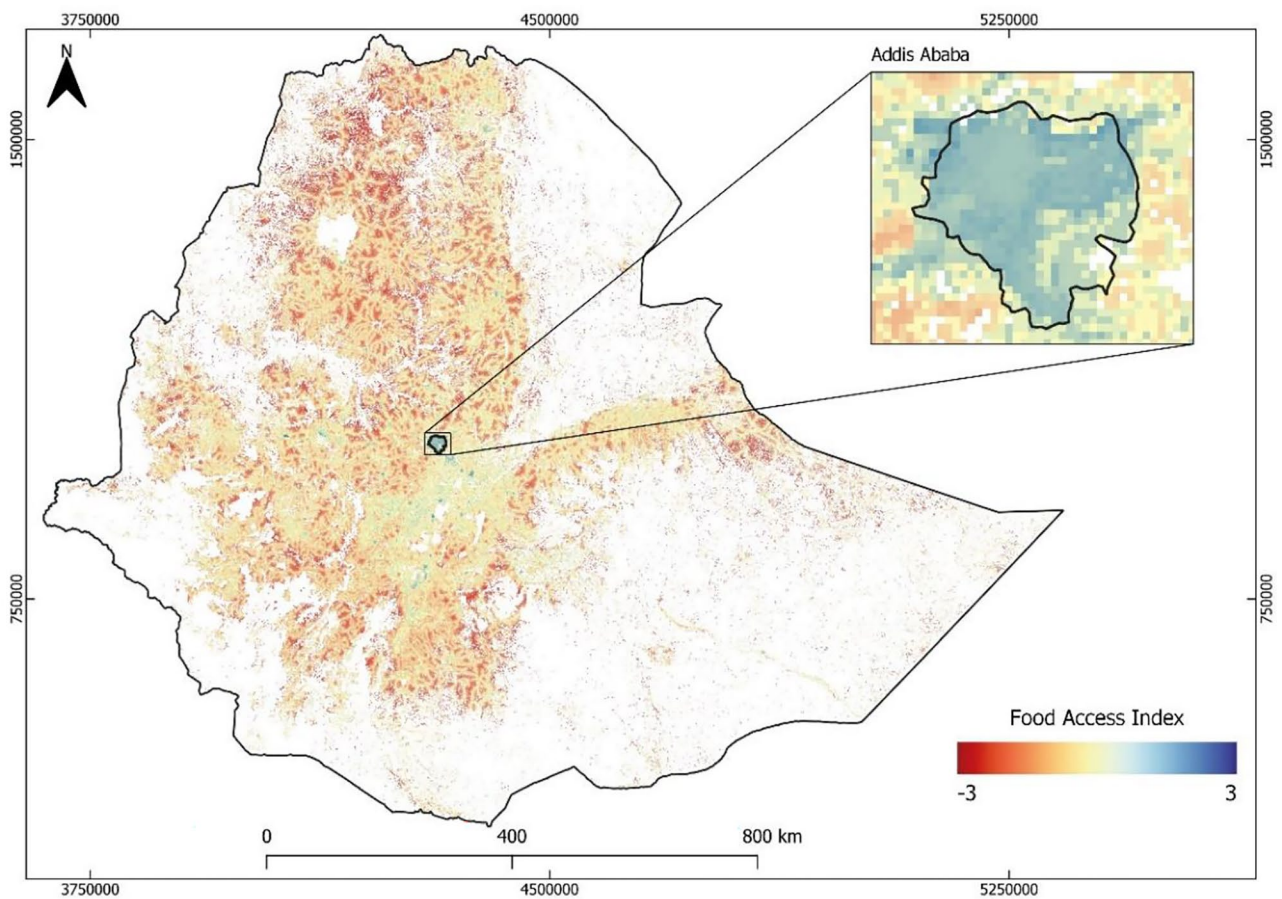


Fig. 6 FAI extrapolated to all of Ethiopia's populated areas

removed since they were not significant at the 95% confidence level in the trained GAM.

The trained GAM using only three geospatial variables explained 57% ($nRMSE = 22.2\%$) of the observed variation in the community-level FAI (Fig. 5) with predicted FAI ranging from -2.5 to 2.9 with a mean of -0.66 and standard deviation of 0.68. When extrapolated to all populated places in Ethiopia (Fig. 6), high FAI values were mostly found in Addis Ababa, and a few neighbouring intermediate and small towns to the south (Mirab Arsi, Misraq Shewa, Alaba, Hadiya, Kembata Tembaro, Sidama, and Silti). Very low FAI values are concentrated in the northwest Amhara and Tigray regions.

The three variables used in the GAM were road proximity, cereal price, and precipitation in the coldest quarter. The PDP plots (Fig. 7) show the relationship of each of these variables to the FAI. The greater the distance from a road, the lower the FAI. FAI was very high (above 1) in areas 0–100 m from a road (class 1), decreased until 500–750 m (class 4), levelled out until 2–2.5 kms (class 8), and reduced again until 3–5 kms (class 10) from a road. There was no training data in areas further than 5 kms from a road (class 11 and 12). FAI did not vary much with precipitation. FAI was highest when cereal prices were lowest and decreased as cereal prices increased until about a value of 7.5.

3.4 Food access disparities across URCA

The FAI per URCA in Fig. 8 showed that all settlements (large, intermediate, and small cities) have a positive mean FAI with uniformly decreasing FAI relative to proximity (within 1, 2, 3 or more hours). Addis Ababa is the only large city in Ethiopia (as per the URCA definition of any city with more than 1 M people) and had the highest FAI with an average FAI of almost 2, while the hinterland areas had the lowest with an average FAI of almost -2. Taken together, all peri-urban areas (< 1 h travel time) contain 66 million people (67%) and have FAI average values less than zero.

4 Discussion

Spatially detailed, nationwide measures of food access are scarce yet there are sufficient sources of relevant data to permit food access to be estimated and mapped at high spatial resolution. We discuss the relevance of publicly available survey and spatial data to food access, how the selected variables and their relationship with food access compare to findings in previous studies, and the suitability and challenges of applying our methodology in other countries.

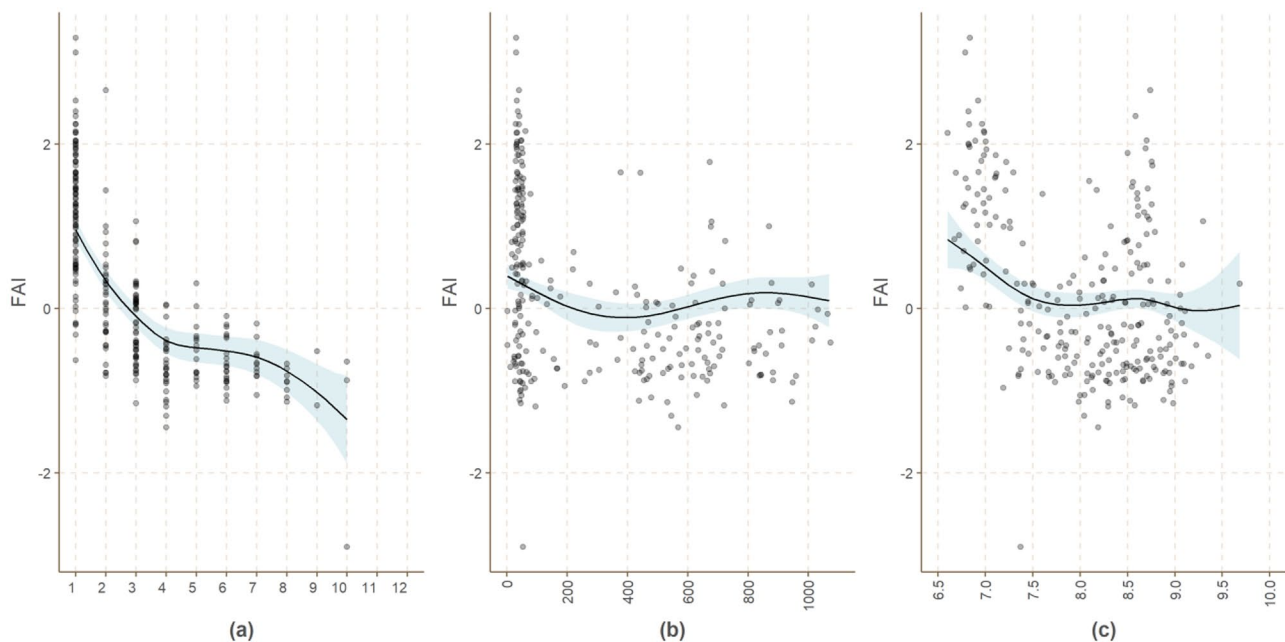


Fig. 7 Partial dependence plots of three significant variables used in GAM: **a** road proximity classes, **b** precipitation in coldest quarter (mm), and **c** cereal price

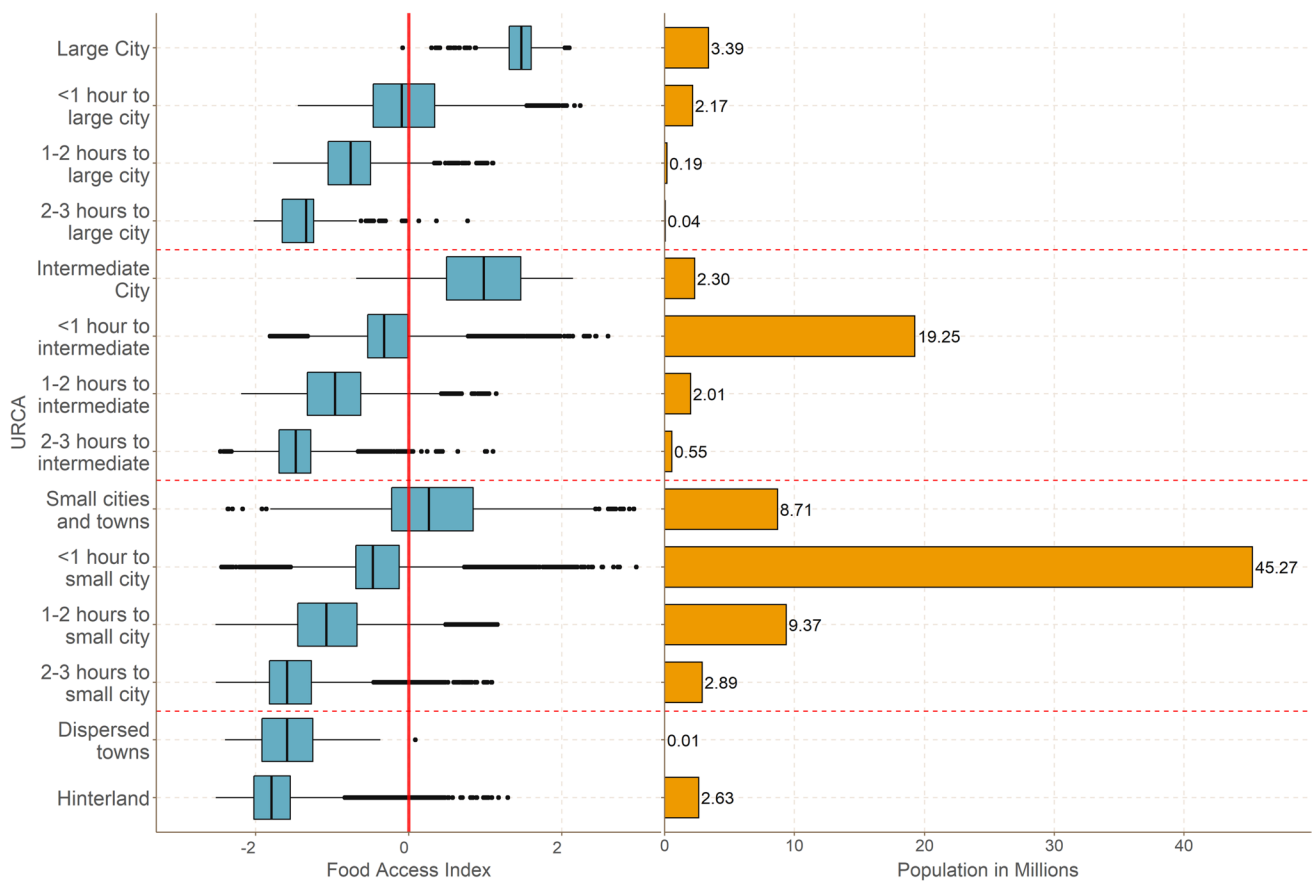


Fig. 8 Boxplot of FAI and corresponding population in millions per URCA

4.1 Relevance of the access-related LSMS variables to independent measures of food access

Most of the 25 variables used to generate the FAI showed significant differences in their means between FI_{access} and FS households. Ownership of a kerosene stove was a component of the FAI but only had a relatively weak contribution to the eventual FAI and explains why it did not have a significant difference between FI_{access} and FS households. Furthermore, total assistance at household level showed a significant difference but this was not the case at the community-level. The differences in means at community level showed that FI_{access} communities had lower means for most distance and fare variables than FS communities. Whilst unexpected, this may reflect that physical access does not completely capture food access since one may be close to food sale locations but not be able to afford the food (Nelson et al., 2019).

The LSMS survey module on Food Insecurity Experience Scale (FIES) can be used to measure the severity of food insecurity. The FIES questions pertained to acute food insecurity experienced in the 7 days prior to the survey. The FIES household responses may represent long-term chronic causes of persistent food insecurity or acute, short-term

causes that manifested themselves in the week prior to the survey. The FAI on the other hand aims to capture the longer-term food access situation. This implies that food access as represented in FIES is not directly comparable to FAI. This highlights the challenge of finding consistent, detailed, countrywide measures of food access, which was one of the motivations for this study.

The FAI methodology can be adapted to represent variability in food access over time. For example, cereal prices vary monthly, while road conditions may vary between dry and wet seasons. These variables, and others that vary over time, can be used to generate a new FAI every month to track temporal changes in food access due to fluctuations in economic and physical factors.

4.2 The contribution of the selected LSMS variables to the FAI

We constructed a food access index (FAI) from LSMS survey data. The first principal component (PC) only accounted for 18.3% of the variance of 25 variables. This is comparable to that of Filmer and Pritchett (2001), whose first PC accounted for 25.6% of the variance for a smaller number of

variables (21) and was still considered robust after interpreting the variables within the first PC. Most of the variables with high absolute value coefficients in the first PC related to economic access. Asset ownership, which had one of the highest coefficients, was as a proxy for wealth in the creation of the DHS relative wealth index (Rutstein & Johnson, 2004). Road quality relates to economic access since urban areas typically receive more infrastructure investments than rural areas (Gertler et al., 2014). Road quality also relates to physical access since better roads improve the transport network and therefore improve access to markets, transport hubs, and other services. Lastly, fares to services can be tied to both economic access (affordability of travel) and physical access (fares are often proportional to distance). This suggests that economic access plays an equal, if not, larger role than physical access in food access in Ethiopia. This is consistent with Hemerijckx et al. (2022) and Lin et al. (2014) who found economic related variables are more important compared to physical or distance related variables.

The positive coefficients of the asset ownership and road quality variables signified that when a community owned more assets or lived in more developed areas, the FAI was higher while the opposite was true when one needed to travel farther or pay more to utilize services. Since the coefficients of proximity and fare to services variables were weaker than those on asset ownership, this suggest that merely being located near urban areas and services does not ensure high food access. This agrees with the assertion that supermarket access (Coveney & O'Dwyer, 2009) and transportation access (Strome et al., 2016) alone are not enough to measure food access.

LSMS is a widely available source of household and community-level data containing relevant information on food security and food access. The selected variables and their empirical relationships to food access may be specific to the economic, social, and physical aspects of food access in Ethiopia. The application of our method to other countries will require country-specific variable screening and selection. As and when additional studies take place, cross-country comparisons of commonly used variables and their weightings may provide additional insight.

4.3 The relevance of selected geospatial variables to the FAI

We started with a large number of geospatial variables and reduced them down to three in three separate pre-processing steps: multicollinearity check, VSURF, and GAM significance. Multicollinearity checks removed 13 bioclimatic variables; this high collinearity has been noted in environmental niche modelling studies (De Marco & Nóbrega, 2018) where bioclimatic data from WorldClim is often used. All crop production variables were removed, most likely because the

original resolution of the MAPSPAM crop production dataset was 10×10 km, which is much coarser than the other geospatial datasets. Out of the 6 remaining bioclimatic variables, 4 were removed by VSURF and 1 was insignificant in the GAM as they were more related to food production and availability than food access. However, these variables may also be important in quantifying food access as our study does not consider cases such as subsistence farming.

The three selected variables from VSURF and GAM were cereal price, road proximity, and precipitation. Cereal prices characterize economic access, specifically the affordability of food (Bondemark, 2020; Breyer & Voss-Andreae, 2013; Lin et al., 2014). Road proximity represents physical access since as one is closer to a road, and without considering paying for fare, it is much easier to utilize the transport infrastructure. Road proximity could also relate to economic access as Wudad et al. (2021) showed that distance to the main road had a significant inverse relationship to household income. Precipitation during the coldest quarter could be a proxy for physical access considering extreme weather such as excessive rain may cause floods affecting the physical transport network which is used for both transporting crops to market and commuting to the market to purchase crops. In the same way, lack of rain (drought) could also cause production obstacles. Adverse weather can also impact economic access since the disruptions to the food system can lead to increased food prices as well as loss of livelihood and income (FAO, 2015; Schmidhuber & Tubiello, 2007).

We interpreted the partial dependence plots (PDPs) from the GAM to help explain the predictions made by the model. The PDP of road proximity classes and cereal prices showed relatively wide confidence bands at the extremes reflecting a lack of data points in areas very far from roads and areas that have very extensive cereal crop areas. LSMS typically conduct surveys in enumeration areas with relatively high populations which are usually near roads. Hence, there are caveats for the interpretation of the FAI in very remote areas which are not represented in the LSMS enumeration areas.

The PDP for road proximity classes revealed the high importance of living within 500 m (up to class 3) to a road for good food access. FAI decreased steeply in communities beyond 2 km from a road. Urban areas have relatively good food access because (1) urban populations have higher economic capacity due to better job opportunities and (2) the quality and quantity of transport network infrastructure and food retail options are both relatively more developed compared to rural areas (Hong et al., 2011) making access to food easier.

The relationship between cereal prices and FAI showed that food access steadily declines as cereal prices increase which is expected based on Gustafson (2013). The PDP also revealed a certain point ($CRP = \sim 7.7$) after which this effect is less evident. This shows that FAI is mostly affected when

cereal prices are low. Our results also reveal that increasing prices beyond ~7.7 no longer substantially impacts FAI. Our cereal prices were an average of the consumer price index (CPI) deflated prices of wheat and maize and not actual prices.

Precipitation during the coldest quarter had a modest contribution to the GAM and interpretation requires some caution and our discussion here is rather speculative. FAI decreased as precipitation increased to 400 mm. The FAO crop calendar of Ethiopia shows that the coldest quarter (December to February) coincides with the time right after the harvest of the main (Meher) season and when crops are being transported to markets. The decrease in FAI could be attributed to transport difficulties brought about by poor road conditions exacerbated by floods (Olana et al., 2018). Second, food access starts decreasing after ~900 mm. In addition to the possible impact of high levels of rainfall on road conditions, this is also the period before the short rains season (Belg) when secondary crops such as legumes are grown. Legumes have relatively low water requirements (250–300 mm). High precipitation and associated oversaturation of soils at the start of the season could also influence production.

Our method uses economic and physical access-related variables both in the constructions of a community-level FAI and its extrapolation to all populated areas. This raises the risk of circularity. We considered this, recognised that it must be properly accounted for, and maintain that the risk is minimal in our final selection of variables. Cereal price did not play a significant role in the community-level FAI (only fruit price did, and even then it was not heavily weighted). Furthermore, the cereal price data used for the extrapolation did not come from LSMS, but from a global food prices database (World Food Programme, 2021). The LSMS variables related to physical access pertain to distance to markets, bus stations and road type, whereas the extrapolation variable is distance to the road network, which again comes from a different source.

4.4 Extrapolation and high-resolution mapping of FAI

The performance of the GAM is comparable to existing food insecurity prediction studies. Lentz et al. (2019) and Deléglise et al. (2022) explained 65% and 47% of the variation in the food consumption score (FCS), respectively, while Westerveld et al. (2021) forecasted mild deterioration of food security status with 68% accuracy. Our results, however, are low compared to Martini et al. (2022) who accounted for 81% of the variation in FCS and 73% of the variation in food-based coping strategies index. There are some important differences to note. Firstly, FCS represents general food insecurity while we concentrate on physical

and economic access to food. Secondly, our method permits a much higher spatial resolution representation of food access, as opposed estimates at specific survey locations or averaged at sub-national administrative boundaries that may not reflect important spatial patterns of food access or food security. A comparison based on method performance may be more appropriate than a comparison against similar measurements.

Our model performs comparatively well to state-of-the-art machine learning methods that explained 56% to 75% of the variation in wealth using geospatial data at similar spatial resolutions (Chi et al., 2022; Jean et al., 2016; Yeh et al., 2020). These studies used between 1,400 to 66,000 samples and uninterpretable geospatial features extracted from satellite imagery using deep learning models. In contrast, our model required substantially less training data (264 samples) and was relatively easy to interpret given only three predictor variables were employed. The PDPs further enhanced the interpretability of the variables. This suggests that our method, with careful selection of the input survey and spatial data, can be used to map food access at high resolution in other countries with similar input data.

4.5 Disparities in food access across URCA

When the FAI predictions were aggregated by URCA, we observed a disparity in food access between urban centres and their catchment areas as well as the farther rural areas. Rural population was found to be an important factor for food deserts, or areas with limited food access (Amin et al., 2021). This finding is also in agreement with Losada-Rojas et al. (2021) who found that urban areas faced lower costs to travel to healthy food retailers compared to rural areas in Indiana, USA. The same study also called for a more equitable distribution of resources to reduce costs in travel towards food sources. The bulk of the population in Ethiopia (65%) is found in the peri-urban (< 1 h travel time) areas of intermediate (19%) and small cities (46%) which is also supported by Cattaneo et al. (2021) who found that intermediate and small cities largely contribute to development in low-income countries because more people are in the catchment areas of such cities. Though these peri-urban areas do not have the lowest food access among all URCA, the FAI values are still relatively low, and the average is below zero. If this finding is corroborated by further studies, then it may suggest that these peri-urban areas should be a focal point for improving food access and the important role of peri-urban areas in increasing food security.

The use of URCA as aggregation units permits a more nuanced assessment of food access across the urban–rural spectrum rather than an urban–rural dichotomy. This would not be possible with food access or food security measures that report at administrative unit, emphasizing the utility of our method to map food access at high spatial resolution.

5 Conclusion

The research aimed to develop and test a transferable method to quantify and map food access, considering both physical and economic dimensions. We successfully constructed a food access index from LSMS survey data and further found that both physical and economic access contributed to food access, but economic access had a stronger influence in the case of Ethiopia. This food access index was also effectively predicted for all populated areas in Ethiopia using a GAM trained only on three geospatial variables. The use of widely available survey and spatial data in the study suggest that the model can be replicated in other countries with suitable training and testing data to pinpoint various demands of inequitable food systems. As an example, we showed that the highly populated peri-urban areas of intermediate and small cities in Ethiopia had low food access suggesting interventions that improve access in these areas could improve food security for a large proportion of the population. Our proof-of-concept model quantifies food access at high spatial detail at a national level using existing open access survey and spatial datasets. This opens opportunities to map economic and physical food access at high spatial resolution in many countries in Sub-Saharan Africa and elsewhere.

Acknowledgements We thank Dr. Kees de Bie and Dr. Ben Sonneveld for their critical comments on the MSc thesis that this article is based on. We also thank Dr. Camilla Bonilla Cedrez for answering queries regarding the cereal price prediction methodology.

Author contributions Conceptualization: Florencio Campomanes V, Andrew Nelson, Michael Marshall; Methodology: Florencio Campomanes V, Andrew Nelson, Michael Marshall; Data curation, formal analysis, writing – original draft: Florencio Campomanes V; Writing – review and editing, supervision: Andrew Nelson, Michael Marshall. All authors have read and approved the final version of the article.

Data availability The authors confirm that the data supporting the findings of this study are available within the article or are in the provided links (for publicly available datasets).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amin, M. D., Badruddoza, S., & McCluskey, J. J. (2021). Predicting access to healthful food retailers with machine learning. *Food Policy*, 99, 101985. <https://doi.org/10.1016/J.FOODPOL.2020.101985>
- Belachew, T., Lindstrom, D., Gebremariam, A., Jira, C., Hattori, M. K., Lachat, C., Huybregts, L., & Kolsteren, P. (2012). Predictors of chronic food insecurity among adolescents in Southwest Ethiopia: A longitudinal study. *BMC Public Health*, 12, 1–11. <https://doi.org/10.1186/1471-2458-12-604/TABLES/6>
- Bergau, S., Loos, T. K., & Sariyev, O. (2022). On- and Off-Farm Diversification and Travel Time to Markets: Linkages to Food Security in Rural Ethiopia. *European Journal of Development Research*, 34, 2543–2560. <https://doi.org/10.1057/S41287-021-00475-0/TABLES/4>
- Bondemark, A. (2020). The relationship between accessibility and price – The case of Swedish food stores. *Journal of Transport Geography*, 82, 102615. <https://doi.org/10.1016/J.JTRANGE.2019.102615>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 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>
- 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>
- Candelise, C., Saccone, D., & Vallino, E. (2021). An empirical assessment of the effects of electricity access on food security. *World Development*, 141, 105390. <https://doi.org/10.1016/J.WORLDDEV.2021.105390>
- Cattaneo, A., Nelson, A., & McMenomy, T. (2021). Global mapping of urban–rural catchment areas reveals unequal access to services. *Proceedings of the National Academy of Sciences*, 118. <https://doi.org/10.1073/PNAS.2011990118>
- Cedrez, C. B., Chamberlin, J., Guo, Z., & Hijmans, R. J. (2020). Spatial variation in fertilizer prices in Sub-Saharan Africa. *PLoS ONE*, 15, e0227764. <https://doi.org/10.1371/JOURNAL.PONE.0227764>
- Central Statistics Agency of Ethiopia. (2020). Ethiopia Socioeconomic Survey (ESS4) 2018–2019. Public Use Dataset. Ref: ETH_2018_ESS_v03_M. <https://doi.org/10.48529/k739-c548>
- Central Statistics Agency of Ethiopia, & World Bank. (2020). Ethiopian Socioeconomic Survey (ESS) 2018/19: SURVEY REPORT. <https://microdata.worldbank.org/index.php/catalog/3823/download/49210>
- Central Statistics Agency of Ethiopia, & World Bank. (2021). Ethiopia Socioeconomic Survey (ESS) baseline basic information document. Addis Ababa, Ethiopia. <https://microdata.worldbank.org/index.php/catalog/3823/download/49208>
- Chen, X. (2017). Take the edge off: A hybrid geographic food access measure. *Applied Geography*, 87, 149–159. <https://doi.org/10.1016/J.APGEOG.2017.07.013>
- Chen, X., & Clark, J. (2016). Measuring Space-Time Access to Food Retailers: A Case of Temporal Access Disparity in Franklin County, Ohio. *Professional Geographer*, 68, 175–188. <https://doi.org/10.1080/00330124.2015.1032876>
- Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences of the United States of America*, 119. <https://doi.org/10.1073/PNAS.2113658119>
- Coveney, J., & O'Dwyer, L. A. (2009). Effects of mobility and location on food access. *Health & Place*, 15, 45–55. <https://doi.org/10.1016/j.healthplace.2008.01.010>

- De Marco, P., & Nóbrega, C. C. (2018). Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS ONE*, *13*, e0202403. <https://doi.org/10.1371/JOURNAL.PONE.0202403>
- Dean, W. R., & Sharkey, J. R. (2011). Rural and Urban Differences in the Associations between Characteristics of the Community Food Environment and Fruit and Vegetable Intake. *Journal of Nutrition Education and Behavior*, *43*, 426–433. <https://doi.org/10.1016/j.jneb.2010.07.001>
- Deléglise, H., Interdonato, R., Bégué, A., Maître d'Hôtel, E., Teisseire, M., & Roche, M. (2022). Food security prediction from heterogeneous data combining machine and deep learning methods. *Expert Systems with Applications*, *190*, 116189. <https://doi.org/10.1016/j.eswa.2021.116189>
- Deller, S., Canto, A., & Brown, L. (2015). Rural poverty, health and food access. *Regional Science Policy & Practice*, *7*, 61–74. <https://doi.org/10.1111/RSP3.12056>
- Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*, 89–121. <https://doi.org/10.1214/SS/1038425655>
- Esayas, B., Simane, B., Teferi, E., Ongoma, V., & Tefera, N. (2018). Trends in extreme climate events over three agroecological zones of Southern Ethiopia. *Advances in Meteorology*, *2018*. <https://doi.org/10.1155/2018/7354157>
- FAO. (2015). Climate change and food security: risks and responses. <https://www.fao.org/3/i5188e/i5188e.pdf>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *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>
- Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). Linear Regression, in: *Biostatistics*. Elsevier, pp. 349–386. <https://doi.org/10.1016/B978-0-12-369492-8.50018-2>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. <http://www.jstor.org/stable/2699986>
- García, X., García-Sierra, M., & Domene, E. (2020). Spatial inequality and its relationship with local food environments: The case of Barcelona. *Applied Geography*, *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 Recognition Letters*, *31*, 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Gertler, P. J., Gonzalez-Navarro, M., Gracner, T., & Rothenberg, A. D. (2014). The role of road quality investments on economic activity and welfare: Evidence from Indonesia's highways. https://sites.bu.edu/neudc/files/2014/10/paper_250.pdf
- Gustafson, D. J. (2013). Rising food costs & global food security: Key issues & relevance for India. *Indian Journal of Medical Research*, *138*, 398.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Prentice Hall.
- Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, *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. *Environmental Research Letters*, *17*, 24016. <https://doi.org/10.1088/1748-9326/AC4D4F>
- Hemerijckx, L. M., Janusz, K., Van Emelen, S., Tumwesigye, S., Davis, J., Lwasa, S., & Van Rompaey, A. (2022). Food accessibility of different socioeconomic groups in sub-Saharan African cities: A mixed-method analysis in Kampala, Uganda. *Food Security*, *14*, 677–694. <https://doi.org/10.1007/S12571-021-01248-7/FIGURES/6>
- Hong, J., Chu, Z., & Wang, Q. (2011). Transport infrastructure and regional economic growth: Evidence from China. *Transportation (amst)*, *38*, 737–752. <https://doi.org/10.1007/S11116-011-9349-6/TABLES/7>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into Principal Components. *Journal of Educational Psychology*, *24*, 417–441. <https://doi.org/10.1037/h0071325>
- International Food Policy Research Institute. (2020). Spatially-disaggregated crop production statistics data in Africa South of the Sahara for 2017. <https://doi.org/10.7910/DVN/FSSKBW>
- 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. *Journal of 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*, *1979*(353), 790–794. <https://doi.org/10.1126/SCIENCE.AAF7894>
- 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. *American Journal of 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. *Applied Geography*, *134*, 102529. <https://doi.org/10.1016/j.apgeog.2021.102529>
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, *72*, 558. <https://doi.org/10.4097/KJA.19087>
- Koh, K., Reno, R., & Hyder, A. (2019). Examining disparities in food accessibility among households in Columbus, Ohio: An agent-based model. *Food Security*, *11*, 317–331. <https://doi.org/10.1007/S12571-019-00900-7/FIGURES/4>
- 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 food-scape 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>
- 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>
- 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.
- Lentz, E. C., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, *122*, 399–409. <https://doi.org/10.1016/j.worlddev.2019.06.008>
- Lin, B. H., Ver Ploeg, M., Kasteridis, P., & Yen, S. T. (2014). The roles of food prices and food access in determining food purchases of low-income households. *Journal of Policy Modelling*, *36*, 938–952. <https://doi.org/10.1016/j.jpolmod.2014.07.002>
- 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>
- Losada-Rojas, L. L., Ke, Y., Pyrialakou, V. D., & Gkritza, K. (2021). Access to healthy food in urban and rural areas: An empirical analysis. *Journal of Transport & Health*, *23*, 101245. <https://doi.org/10.1016/j.jth.2021.101245>

- Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18, 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., & Omodei, E. (2022). Machine learning can guide food security efforts when primary data are not available. *Nat Food*, 3, 716–728. <https://doi.org/10.1038/s43016-022-00587-8>
- Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, 28, 268–280. <https://doi.org/10.1177/002224379102800302>
- Mohamed, A. A. (2017). Food security situation in Ethiopia: A review study. *International Journal of Health Economics and Policy*, 2, 86–96.
- Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019). A suite of global accessibility indicators. *Scientific Data*, 6, 1–9. <https://doi.org/10.1038/s41597-019-0265-5>
- Nguyen, T., Mai, H. P. T., van den Berg, M., Thanh, T. H. T., & Béné, C. (2021). Interactions between food environment and (Un)healthy consumption: Evidence along a rural-urban transect in Viet Nam. *Agriculture*, 11, 789. <https://doi.org/10.3390/AGRICULTURE11080789>
- 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. *International Journal of Environmental Research and 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: Analyzing crop production, availability, and market functions for 2016/2017 and estimations for 2017/2018. https://www.usaid.gov/sites/default/files/2022-05/Crop_Availability_and_Market_Analysis_in_Ethiopia_05.09.18_508_Compliant_1.pdf
- OpenStreetMap contributors. (2022). Planet dump [Data file from 13/02/2022] [WWW Document]. Accessed date April 9, 2023, from <https://www.openstreetmap.org>
- Ploeg, M. V., Dutko, P., & Breneman, V. (2015). Measuring food access and food deserts for policy purposes. *Applied Economic Perspectives and Policy*, 37, 205–225. <https://doi.org/10.1093/AEPP/PPU035>
- Pozzi, F., & Robinson, T. (2008). Accessibility mapping in the horn of Africa: Applications for livestock policy, IGAD LPI Working Paper 11–08.
- Rutstein, S. O., & Johnson, K. (2004). The DHS Wealth Index, DHS comparative reports No. 6. Calverton, Maryland: ORC Macro.
- Salima, W., Manja, L. P., Chiwaula, L. S., & Chirwa, G. C. (2023). The impact of credit access on household food security in Malawi. *Journal of Agriculture and Food Research*, 11, 100490. <https://doi.org/10.1016/J.JAFR.2022.100490>
- Schmidhuber, J., & Tubiello, F. N. (2007). Global food security under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19703–19708. <https://doi.org/10.1073/PNAS.0701976104>
- Sharkey, J. R., & Horel, S. (2008). Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *Journal of Nutrition*, 138, 620–627. <https://doi.org/10.1093/jn/138.3.620>
- Strome, S., Johns, T., Scicchitano, M. J., & Shelnett, K. (2016). Elements of access: The effects of food outlet proximity, transportation, and realized access on fresh fruit and vegetable consumption in food deserts. *Community Health Equity Research & Policy*, 37, 61–70. <https://doi.org/10.1177/0272684X16685252>
- Wang, H., Liu, C., Fan, H., & Tian, X. (2017). Rising food accessibility contributed to the increasing dietary diversity in rural and urban China. *Asia Pacific Journal of Clinical Nutrition*, 26, 738–747. <https://doi.org/10.6133/apjcn.052016.03>
- Wantchekon, L., & Riaz, Z. (2019). Mobile technology and food access. *World Development*, 117, 344–356. <https://doi.org/10.1016/J.WORLDDEV.2019.01.006>
- Westerveld, J. J., van den Homberg, M. J., Nobre, G. G., van den Berg, D. L., Teklesadik, A. D., & Stuit, S. M. (2021). Forecasting transitions in the state of food security with machine learning using transferable features. *Science of the Total Environment*, 786, 1–15. <https://doi.org/10.1016/j.scitotenv.2021.147366>
- Wood, S. N. (2017). Generalized additive models: An introduction with R, second edition. Generalized Additive Models: An Introduction with R, Second Edition 1–476. <https://doi.org/10.1201/9781315370279>
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73, 3–36. <https://doi.org/10.1111/J.1467-9868.2010.00749.X>
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65, 95–114. <https://doi.org/10.1111/1467-9868.00374>
- World Food Programme. (2021). Global food prices database. <https://data.humdata.org/dataset/wfp-food-prices>
- WorldPop, Center for International Earth Science Information Network, Columbia University. (2018). Ethiopia population counts 2018 UN adjusted. <https://doi.org/10.5258/SOTON/WP00675>
- Wudad, A., Naser, S., & Lameso, L. (2021). The impact of improved road networks on marketing of vegetables and households' income in Dedo district, Oromia regional state, Ethiopia. *Heliyon*, 7. <https://doi.org/10.1016/J.HELIYON.2021.E08173>
- Yamaguchi, M., Takahashi, K., Hanazato, M., Suzuki, N., Kondo, K., & Kondo, N. (2019). Comparison of objective and perceived access to food stores associated with intake frequencies of vegetables/fruits and meat/fish among community-dwelling older Japanese. *International Journal of Environmental Research and Public Health*, 16, 772. <https://doi.org/10.3390/IJERPH16050772>
- 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. *Nature Communications*, 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). <https://doi.org/10.5281/zenodo.5571936>



Florencio Campomanes V is a PhD Student at the University of Twente, faculty of Geo-Information Science and Earth Observation (ITC) where he also obtained his MSc in Geoinformation Science and Earth Observation (natural resources management). His research interests focus on the integration of remote sensing, geospatial data, and artificial intelligence in the field of urban development and food access.



Dr. Michael Marshall is an assistant professor at the University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Department of Natural Resources. Michael seeks to improve our understanding of crop carbon and water exchange from field to global scales. Michael answers his research questions with models that capture the spatio-temporal dynamics of agricultural production

(area \times yield) and water use (evapotranspiration). The models are: (i) implemented in a Geographic Information System; (ii) based on principles of radiation-use efficiency and machine-learning; and (iii) driven by Earth observation, climate reanalysis, and other geospatial information. Ultimately, the results of his research inform food security analysis and water resources management. He currently leads two European Space Agency projects: *EO4NUTRI—Earth Observation for estimating and predicting crop NUTRIents* and *HyRELIEF—Enhancing ECOSTRESS drought monitoring with hyperspectral narrowbands*.



Dr. Andy Nelson is full professor in Spatial Agriculture and Food Security at the University of Twente, faculty of Geo-Information Science and Earth Observation (ITC), department of Natural Resources. His academic work uses Earth observation and spatial information for improved food security. His academic work – with students and colleagues at ITC and around the world - concerns (i) where, when, and how is food produced, (ii) how much is produced, what are the risks to production, and how can these risks be mitigated, and (iii)

how do food and people move from A to B and what happens in the event of disruptions or changes to our food transport network infrastructure. He serves on the Advisory Board of Food Security and is also a Senior Editor for Food Security.