Food Security (2021) 13:55–70
https://doi.org/10.1007/s12571-020-01132-w

ORIGINAL PAPER

A nonparametric analysis of household-level food insecurity and its determinant factors: exploratory study in Ethiopia and Nigeria

Maryia Bakhtsiyarava[1] · Tim G. Williams[2] · Andrew Verdin[3] · Seth D. Guikema[2]

Received: 17 June 2020 / Accepted: 7 December 2020 / Published online: 14 January 2021

Abstract

Given the fundamental importance of food to human well-being, understanding food insecurity is crucial for sustainable development. However, due to the complex nature of food insecurity, traditional linear methods of empirical analysis may mask critical relationships between food insecurity and demographic, agricultural, and environmental factors. Here we show, using two years of household-level survey data from Ethiopia and Nigeria, that nonparametric regression ("random forest", in this study) enables enhanced insight into the factors associated with self-reported food security and household dietary diversity score. We observe nonlinearities and thresholds in the relationships between the measures of food security, livestock ownership, and climatic conditions. The threshold-based relationships suggest that policies aimed at increasing agricultural productivity (e.g., livestock holdings) may only be beneficial up to an extent. While it is intuitive that some level of diminishing returns will exist, our nonparametric analysis could be used as a first step to discern the levels to which policies may be beneficial. Additionally, our results indicate that the random forest (and perhaps nonparametric regression and classification methods more generally) may be especially well-positioned to uncover nuances in these relationships in years with suboptimal climatic conditions (such as during the 2015 drought in Ethiopia). Ultimately, we argue that nonparametric approaches, when informed by existing theory, provide an insightful complement to inform the analysis of agricultural and development policy.

Keywords Food security · Agriculture · Nonparametric regression · Statistical inference

1 Introduction

Food is essential to human well-being, and food shortages deteriorate health, exacerbate poverty, and contribute to political and socioeconomic instability. Despite substantial progress in reducing the world's malnourished population over the last decades, food insecurity is still a pervasive problem (Myers et al. 2017; Schmidhuber and Tubiello 2007; Wheeler and Von Braun 2013). Given the compounding effects of a growing population and the projected impacts of climate change (Schlenker and Lobell 2010; Schlenker and Roberts 2009), this trend presents a challenge for achieving the Sustainable Development Goal # 2, which is aimed at eliminating hunger and malnutrition by 2030 (UNDP 2015).

Various dimensions of food insecurity such as hunger, malnutrition, and their determinants are understood via the concept of food security, which is defined as a situation when people have constant and reliable access to the food that fits their needs and preferences for a healthy and active lifestyle (FAO 1996). Embedded in this concept are four so-called pillars, or determinants of food security: availability, access, utilization, and stability (FAO 1996). Availability refers to the sufficient quantity of food on the market and reflects aggregate crop production/imports. Access refers to the ability of households or individuals to obtain food. Utilization reflects

✉ Maryia Bakhtsiyarava
mariab@berkeley.edu

Tim G. Williams
tgw@umich.edu

Andrew Verdin
averdin@umn.edu

Seth D. Guikema
sguikema@umich.edu

[1] Institute of Urban and Regional Development, University of California - Berkeley, Berkeley, CA 94720, USA

[2] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, USA

[3] Minnesota Population Center, University of Minnesota, Minneapolis, MN 55455, USA

both differences in food allocation within households and the biological process of bodies absorbing and metabolizing nutrients from the food (Jones et al. 2013). Stability refers to the consistency of food availability and access across time and space. Food insecurity occurs when one or more of the four pillars are compromised.

In countries where a large proportion of the population is engaged in agriculture, the performance of the agricultural sector is of particular importance for household- and national-level food security. Therefore, in theory, agriculture as a livelihood activity has a large potential to influence food security, especially for small-scale farmers. However, the majority of empirical research has not been able to corroborate the link between agriculture and food and nutritional security: even though various agricultural strategies (such as increasing agricultural production, cultivating more crops, diversifying livestock herd, etc.) can increase household food production, they do not necessarily improve household nutrition (Carletto et al. 2015; Webb and Kennedy 2014). In particular, empirical studies reveal the critical importance of markets, access to off-farm income, and agro-ecological context in mediating agriculture-food security linkages for smallholder households (Frelat et al. 2016; Jones 2017). These findings point to the immense complexity of food security. The complexity of food security warrants diverse methodological approaches to understand it (Müller et al. 2020). In fact, a lack of conclusive evidence between specific agricultural activities and food security outcomes may stem, in part, from the methodological approaches to analyzing these relationships (Webb and Kennedy 2014). Empirical research on food security is largely dominated by parametric statistical methods, while nonparametric methods to analyze food security remain somewhat underutilized, despite earlier studies (Barrett and Dorosh 1996; Deaton 1989) and advances in data availability and computing power.

This study, set in Ethiopia and Nigeria, contributes to the literature by analyzing and comparing associations between household demographic, agricultural, and climatic factors and food security outcomes as they emerge from the application of a nonparametric method (random forest) and traditional parametric methods (linear and logistic regressions). The general idea is that in using different analytic techniques, the analysis provides a more nuanced perspective of food security that is less dependent on assumptions about the underlying data. Specifically, we focus on the relationships between livestock ownership, the number of crops cultivated by the household, and rainfall and temperature conditions as they relate to two household food security outcomes: self-reported food security status and household dietary diversity score. From a policy perspective, this type of analysis could be used to examine limits and thresholds in food security relationships, and how these vary in different contexts. One can also build on these techniques to identify populations most vulnerable to food insecurity.

Our specific goals are to examine (1) how measures of food insecurity are associated with agricultural, climatic, and demographic factors according to parametric and nonparametric methods, (2) whether these associations are nonlinear, and (3) whether the observed relationships differ between Ethiopia and Nigeria, two sub-Saharan African countries with different environmental conditions, pervasive food insecurity, and a large dependence on agriculture for livelihoods.

2 Background

2.1 Alternative data and methods for empirical food security research

Recent improvements in computing power, easily accessible satellite data, as well as publicly available gridded datasets (Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is one example) have been harnessed to address the complexity of food security and provide auxiliary information to power food security research (Grace et al. 2012, 2014; Kugler et al. 2019; Lentz et al. 2019; Lobell et al. 2020). Relevant to our work, a recent study by Lentz et al. (2019) spatially and temporally integrated a suite of data on rainfall, livelihood zones, market price data, and sociodemographic characteristics to create a predictive model of food security for Malawi. Evaluation of the predictive accuracy of the model used in the study suggested that data-driven approaches to identifying hotspots of food insecurity should be used to improve food security crisis response.

Another group of studies has used machine learning methods to integrate survey and satellite data to study social outcomes (Frelat et al. 2016; Jean et al. 2016; Yeh et al. 2020). For example, Frelat et al. (2016) used artificial neural networks to examine relationships between household-level food availability and socio-demographic and environmental characteristics in a dataset comprising more than 13,000 households. They found that land area, livestock holdings, and household size explained a substantial amount of the variability in food availability, and that these relationships were both nonlinear and strongly affected by market access and environmental constraints.

Despite these studies, there have been few efforts to apply nonparametric techniques to understand the relationships between food security outcomes and their determinants. Traditionally, studies that evaluate the associations between food security and determinant factors generally assume linear relationships (Headey and Hoddinott 2016; Sibhatu et al. 2015; Sibhatu and Qaim 2017). While providing important insight into potential linkages, such studies assume that the associations between, for example, agricultural production and food shortages are constant throughout the range of observed values of production volume. Conceptually, this can be a tenuous assumption, as various factors are theorized to exhibit nonlinear relationships with food security. For example, a U-shaped relationship between precipitation and crop yields may occur in contexts where too little or too much rain is associated with failed crops (Husak and Grace 2016). Quadratic and cubic transformations of covariates have been used to investigate nonlinear relationships and thresholds within a parametric regression framework (Nawrotzki et al. 2017); however, parametric methods do not allow such relationships to emerge organically. Building on the theoretical understanding of the determinants of food security that emerged from previous research and the data-driven approaches demonstrated in the studies discussed above, we apply nonparametric methods in combination with standard regression techniques to improve our understanding of the topic.

2.2 Nonparametric methods

2.2.1 How nonparametric methods work: overview

Nonparametric regression methods (NPM) allow for flexibility in the predictor and, as opposed to quadratic or cubic parametric approaches, do not require the relationship between the predictor and outcome variables to be presupposed, potentially allowing for a more nuanced understanding of the observed associations (Kuku et al. 2011). Nonparametric regression techniques also often do not require such strict assumptions about the distribution of the data (e.g., normality) and have proven to better predict data not used to train the model (Shortridge et al. 2015), so subsequent descriptive inferences can be more reliable. Additionally, nonparametric regression is often as easy to specify in common statistical programming languages (e.g., R) as its parametric counterpart. Therefore, given the persistence of global food insecurity and the complexity of food systems, nonparametric analysis has the potential to expand the food insecurity discussion through more reliable predictions and nuanced insight.

2.2.2 Nonparametric methods in social science

Nonparametric methods (NPM) have been used previously to investigate social outcomes. In social science research, NPMs have been used for density estimations and graphic presentation of the results. For example, Deaton (1989) used density kernel smoothing to describe the impact of rice prices on household income distribution in Thailand. The study concluded that some of the observed patterns, such as a nonlinear relationship between an increase in rice prices and household income, with price hikes mostly benefitting households in the middle of the income distribution and not rich and/or poor households, would not have been revealed by standard econometric techniques (Deaton 1989). Specifically, Deaton (1989) pointed out that the standard econometric techniques would not have been able to uncover relationships at the tails of the income distribution. In a similar vein, Barrett and Dorosh (1996) used kernel density smoothing to analyze the effect of grain prices on farmers' welfare in Madagascar. The study extended the nonparametric approach introduced in Deaton (1989) by, among other things, estimating conditional expectations (that can be thought of as probabilities) of households' purchasing and selling of food relative to their income and land holdings (Barrett and Dorosh 1996). Another notable contribution of their study is the bootstrapping approach used to estimate confidence intervals that does not rely on asymptotic assumptions about the data. More specific to the current study on food security, Kuku et al. (2011) conducted a parametric and nonparametric assessment of the relationship between child obesity and household food security. They supplemented the use of linear parametric regressions with the nonparametric Local Weighted Scatterplot Smoothing (LOWESS). LOWESS is another nonparametric technique that works like a local weighted regression and builds on kernel density smoothing. The parametric results from their study showed no relationship between household food security and child obesity. LOWESS results, however, revealed a more complex and nuanced relationship, with the risk of child obesity being highly nonlinear and dependent on demographic and socioeconomic factors.

A common thread in the studies described above is the recognition of the capability of NPMs to uncover unexpected, nonlinear, and complex relationships that cannot so easily be uncovered with parametric models. In allowing the data to speak for themselves, NPMs do justice to the richness of the data collected in the surveys without imposing assumptions about data distribution. Still, the application of nonparametric methodological approaches to understand the underlying relationships in food security research has been rare. An advantage of the nonparametric approach we undertake here (described in more detail in the Methods section) is that it allows us to evaluate the relationship between a measure of food security and a determinant factor of interest (see following section) while accounting for the other determinant factors. As such, we take advantage of the nonparametric approach to uncover the underlying relationships while controlling for the influence of other factors on food security. We argue that complementing traditional methods with nonparametric methods, grounded in existing theory, will improve our understanding of the current complex food security situation and what can be done to improve it.

2.3 Determinants of food security

In this study we are primarily interested in the agricultural and environmental determinants of food security. Guided by previous research, we focus on analyzing relationships between the following agricultural and environmental factors and food security outcomes. *Yield (kg/ha)* is a measure of agricultural productivity, and we consider yields to reflect food availability at the household level, especially if a household is engaged in subsistence farming; in the case of farmers who are net-sellers, yield is related to agricultural income. *The number of unique crop varieties* cultivated by the household may reflect dietary diversity, with different crops having different dietary value (Shively and Sununtnasuk 2015; Sibhatu et al. 2015). Crops also have varying tolerance to pests and weather variability, which suggests that cultivating more crop varieties may be a strategy to reduce risk. Similar to the number of crops, *the number of livestock* has also shown a relationship with food and nutritional security outcomes (Abay and Hirvonen 2017; Frelat et al. 2016; Hoddinott et al. 2015). We include measures reflecting household ownership of large ruminants (cows and oxen) and small ruminants (goats and sheep). These livestock are valuable household assets because they are as a source of draft power and manure for land cultivation, and are a significant source of milk and meat (Devendra 2005; Sansoucy 1995; Workneh et al. 2003). Finally, we also account for *off-farm income* as it has been shown to play a substantial role in household food security, nutrition, and general welfare (Babatunde and Qaim 2010; Reardon 1997).

Agriculture in Ethiopia and Nigeria is primarily rainfed, making its performance highly sensitive to climatic conditions. Ethiopia in particular has a long history of droughts that are detrimental to agriculture (Devereux and Sussex 2000; Funk et al. 2012), and the country recently experienced an extensive drought in 2015 (Philip et al. 2018). Therefore, this study also considers temperature and rainfall conditions that are relevant to the performance of agriculture (described in the next section).

3 Materials and methods

3.1 Household survey data

We rely on the 2013-2014 and 2015-2016 rounds of the Ethiopia Socioeconomic Survey and 2012-2013 and 2015-2016 rounds of the Nigeria General Household Survey, conducted in cooperation with the World Bank's Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) (World Bank 2019). These georeferenced, large-scale surveys collect detailed data on demographic characteristics, health, agriculture, time use and labor, food security and shocks, and banking and credit. The surveys are spatially referenced at the level of Enumeration Areas (EA), which roughly represent the village level. Because of our focus on households engaged in agriculture, we restrict the sample to rural households. Our final sample includes 3055 and 2954 households for Nigeria LSMS in 2013 and 2016, and 2960 and 2809 households for Ethiopia LSMS in 2014 and 2016, respectively. The households were interviewed during the months of February-April during each survey, after they had finished planting and started harvesting their crops. While we constructed the majority of the variables used for the analysis from the raw LSMS microdata distributed directly by the World Bank, we also used several variables produced from the same LSMS data and distributed by the Evans School Policy Analysis and Research Group at the University of Washington (EPAR) (EPAR 2019). EPAR uses the World Bank's LSMS data to create agricultural development indicators that are comparable across countries. We used EPAR measures of farm area, yield, and off-farm income to ensure consistency and comparability in units between Nigeria and Ethiopia.

3.1.1 Outcome measures

We employ two outcome variables. The first is a subjective, binary indicator of food security, in which households self-reported experiencing food shortage(s) (food insecure) or no food shortages (food secure) within the past year. This reporting was done at the household level, and a food shortage represents not having enough food to feed the household at some point in the past 12 months. The second measure is a household dietary diversity score (HDDS), measured by adding up consumption of 12 different food groups by the household in the last week (Swindale and Bilinsky 2006). HDDS can take on values from 0 to 12, with higher score reflecting a more diverse diet. The two measures we rely on describe food security from different angles: the first measure reflects subjective, self-reported experiences with food shortages, whereas HDDS is more of an objective measure that reflects diet quality.

3.1.2 Sociodemographic variables

In line with previous research, we include the following sociodemographic variables that may impact household food security: sex and age of household head, household size, dependency ratio (the ratio of the number of household members 65 years and older and 15 years and younger to the number of household members of working age (15-65)), and off-farm income. These have shown associations with household food security in previous research (Abafita and Kim 2014; Beyene and Muche 2010; Bogale and Shimelis 2009; Demeke et al. 2004; Feleke et al. 2005) and often correlate with household access to resources and poverty, all of which can impact food security outcomes.

3.1.3 Agricultural variables

As described above, this study focuses on the following agricultural characteristics that have been shown to impact food security outcomes: yield (kg/ha), the number of crop varieties cultivated by the household, and the number of large ruminants (cows and oxen) and small ruminants (goats and sheep) owned by the household. We also account for household farm size.

3.2 Climate data and measures

We construct a range of variables to account for recent and long-term weather trends. The source of rainfall data is the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). CHIRPS rainfall data are available at a high level of spatial detail (0.05°) from 1981 up to date (Funk et al. 2015). For temperature, we used the Climate Hazards Center Infrared Temperature with Stations (CHIRTS), which is a new, high-resolution global temperature dataset produced by the same group as CHIRPS (Funk et al. 2019). A suite of sources are used to produce CHIRPS and CHIRTS data, including satellite estimates and weather data from stations (Funk et al. 2015, 2019).

We construct the following measures to describe climatic conditions in the study area. First, we compute a measure of total rainfall during the last 12 months before each survey for a 10-km buffer for every EA (village). Survey questions on food security were asked in February–April of every survey year (2014 and 2016 for Ethiopia; 2013 and 2016 for Nigeria). By that time, farmers had planted their fields and commenced harvest. Therefore, a 12-month window represents February 2015–January 2016 for Ethiopia and Nigeria 2016 LSMS, for example. Second, we compute average maximum monthly temperature for the same 12 months before the survey. These rainfall and temperature measures for the last 12 months reflect growing conditions for the crops. Finally, we also compute temperature and rainfall anomaly z-scores relative to the long-term (1981–2016) average rainfall and long-term (1983–2016) average maximum temperature. Z-scores describe deviations of rainfall and temperature from the long-term normal, and thus allow us to investigate how food security outcomes are associated with actual changes in climatic conditions. Detailed variable description and summary statistics can be found in Table 1.

3.3 Methods

3.3.1 Nonparametric regression and classification: random forest

We experimented with several common nonparametric methods, all of which are simple to implement in the R environment for statistical computing (R Core Team 2019) and other programming languages. However, we retained only the random forest for the inferential analysis as it provided the highest predictive accuracy of the methods we tested. Implementation details and justification for the selection of random forests are described in the Supplementary Information.

A *random forest*, originally developed by Breiman (2001), consists of an ensemble of regression or classification trees. Each tree is built by recursively "splitting" the dataset at points within the domain of the independent variables (e.g., splitting the dataset into two groups representing $X_3 < 4.6$ and $X_3 \geq 4.6$). For classification, the splitting attempts to most accurately partition the response variable into its classes (in this case, instances of food security and insecurity). In each terminal node (or "leaf") of a classification tree, the model predicts the most common class. For regression applications, the splitting attempts to reduce a measure of the variability of the response in each leaf (in this case, HDDS values). The model then predicts the average value within each leaf. Random forests consist of a large number of trees (e.g., 500), each of which is independently trained using a bootstrapped sample of the original data. Additionally, at each splitting point, only a random subset of the independent variables is considered. These properties help to reduce the correlation between the trees, decreasing the overall model variance (Deng et al. 2011). Each tree is used to provide a prediction for an input vector **X**. For classification, the final prediction for the overall random forest can be given either by the most popular classification or stated as a probability, $P_{\mathbf{X},i}$, representing the probability that the vector **X** was assigned to class i. For regression, the final prediction is the average prediction over all trees. Because of their ensemble nature and the fact that each tree is only fit to a subset of the covariates, random forests are less prone to overfitting than regular tree-based methods. Consequently, they have proven widely popular and generally perform well at prediction.

To explore the associations captured by the models we use partial dependence plots (PDPs). PDPs are visualizations of the association between each independent variable (X_i) and the measure of food insecurity (Y), accounting for the variability in all other covariates. Each point (x, y) on a partial dependence plot represents the average prediction made by the model (y value) if every instance of X_i is set to x, keeping all other independent variables (X_{-i}) at their original values. A PDP for a linear regression model would show a straight line representing the regression coefficient (β).

We estimate two model specifications for the random forest and the logistic and linear regression (discussed in the next section). Model 1 includes the following independent variables: age and sex of household head, household size, household dependency ratio, household floor type, off-farm income, farm size, yield, number of crops

Table 1 Variable description and summary statistics for the 2016 LSMS samples for Ethiopia and Nigeria

Variable name	Variable description	Ethiopia		Nigeria	
		Mean	SD	Mean	SD
Outcome variables					
Self-reported household food security status	1- food secure: a household did not report experiencing not having enough food in the last 12 months; 0 - food insecure (% food secure households)	71.81		81.89	
Household Dietary Diversity Score (HDDS)	Count of food groups consumed by the household in the last 7 days (0–12)	5.81	1.82	8.20	1.88
Independent variables					
Sociodemographic					
Age of household head	Age of household head (years)	46.70	14.26	51.60	13.87
Sex of household head	Sex of household head (% female-headed households) [UW]	22.79		16.89	
Household size	Total number of household members [UW]	6.34	2.49	7.74	3.57
Dependency ratio	The ratio of the number of household members younger than 15 and older than 65 to the number of members between the ages of 15–64	1.17	0.90	1.05	0.89
Floor	Finished floor in the household's dwelling (% households)	3.60		60.73	
Non-agricultural income	Income from wage employment in all non-agricultural activities (2016 USD) [UW]	182.00	5556	541.70	6520
Agricultural					
Farm size	Land size, all cultivated plots (ha) [UW]	1.21	8.30	1.04	2.61
Yield	Yield of all major crops (kg/ha) [UW]	71,563	351,576	112,321	547,731
Number of crops	Total number of crop varieties cultivated by the household	5.61	3.97	2.78	1.90
Number of livestock	Total number of cows, oxen, goats, and sheep owned by the household	9.71	22.44	4.96	13.39
Climatic conditions					
Rainfall	Total rainfall during the 12 months before the survey (mm)[2]	985	482	1358	534
Temperature	Average maximum monthly temperature during the last 12 months before the survey (degree C)	27.73	3.25	33.0	1.42
Rainfall anomaly z-score	Z-score of the total rainfall 12 months before the survey relative to 1981–2016	−1.06	1.57	−0.72	0.76
Temperature anomaly z-score	Z-score of the average max temperature 12 months before the survey relative to 1983–2016	1.88	0.51	1.17	0.44
N (households)		2809		2954	

[1] UW indicates a variable was sourced from the LSMS data distributed by the Evans School Policy Analysis & Research Group at the University of Washington (EPAR 2019); otherwise it was constructed from the raw World Bank LSMS microdata

[2] In Nigeria and Ethiopia, household sections of the survey were fielded in February–April, by which time planting ends and harvesting starts. All climate variables were computed for the 12 months before the start of the household section of LSMS (for example, February 2015–January 2016 for Ethiopia and Nigeria 2016 LSMS)

cultivated, number of livestock, total rainfall over the last 12 months before the survey, and average maximum temperature over the last 12 months. Model 2 includes the same variables as Model 1, except the 12-months climate variables are replaced by rainfall and temperature anomalies' z-scores; Model 2 also controls for long-term average maximum temperature (1983–2016) and rainfall (1981–2016). Model 1 and Model 2 include fixed effects for regions (Ethiopia) and states (Nigeria).

3.3.2 Logistic and linear regressions

We compare the random forest results alongside those of a logistic regression (for the binary outcome of household food security status) and linear regression (for the continuous measure of HDDS). Logistic (and probit) regressions are commonly employed in studies of food security where the outcome is dichotomous (Bogale and Shimelis 2009; Demeke et al. 2011; Sam et al. 2019). Our logistic model takes the form $logit(Y) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i$, where Y is the dichotomous outcome variable, X_i are the independent variables, and β_i are the estimated regression coefficients. β_i represent log odds, which were converted to probabilities for comparison with the random forest results. The models include regions (for Ethiopia) and states (for Nigeria) as fixed effects to account for unobserved heterogeneity in the linear and logistic regressions. Standard errors are clustered at the EA level because it is the level at which climate data were aggregated. We computed Variance Inflation Factors (VIF) to ensure no significant multicollinearity among the independent variables.

3.3.3 Statistical significance and variable importance

Parametric methods generally calculate a p-value for each independent variable, which represents the probability that the regression coefficient is statistically different from zero. Nonparametric methods, however, do not automatically generate an equivalent metric. This may be seen as a limitation by many researchers who are accustomed to using statistical tests to inform their interpretation. However, notions of "variable importance" can be derived from the models (Breiman 2001), which describe the model's assessment of the magnitude of influence exerted by each independent variable on the dependent variable. Additionally, the vertical range of the PDP can be used as another indication of the relative influence of a variable on the outcome (Shortridge et al. 2015). In both of these approaches, there is no "importance threshold" (e.g., $p < 0.05$), below which variables are discarded from the analysis. This is because, given the nature of nonparametric models, different variables can have different degrees of influence over their domain. PDPs can be used to assess this level of influence and make more qualitative judgments about the relative effect of different factors.

4 Results

We present the results by describing patterns that emerge from both logistic/linear regressions and nonparametric approaches. We present results in the form of partial dependence plots, which depict associations between each covariate and the food security outcomes, accounting for the variability in all other covariates. We plot these for both the logistic/linear regressions and the random forest. The first general pattern we observe is that the relationships between the two measures of household food security and their determinants vary from one year to another, not only in magnitude but also in the direction. While this finding is evident from the regression results alone, the random forest provides a way to explore these patterns in more detail.

4.1 Livestock

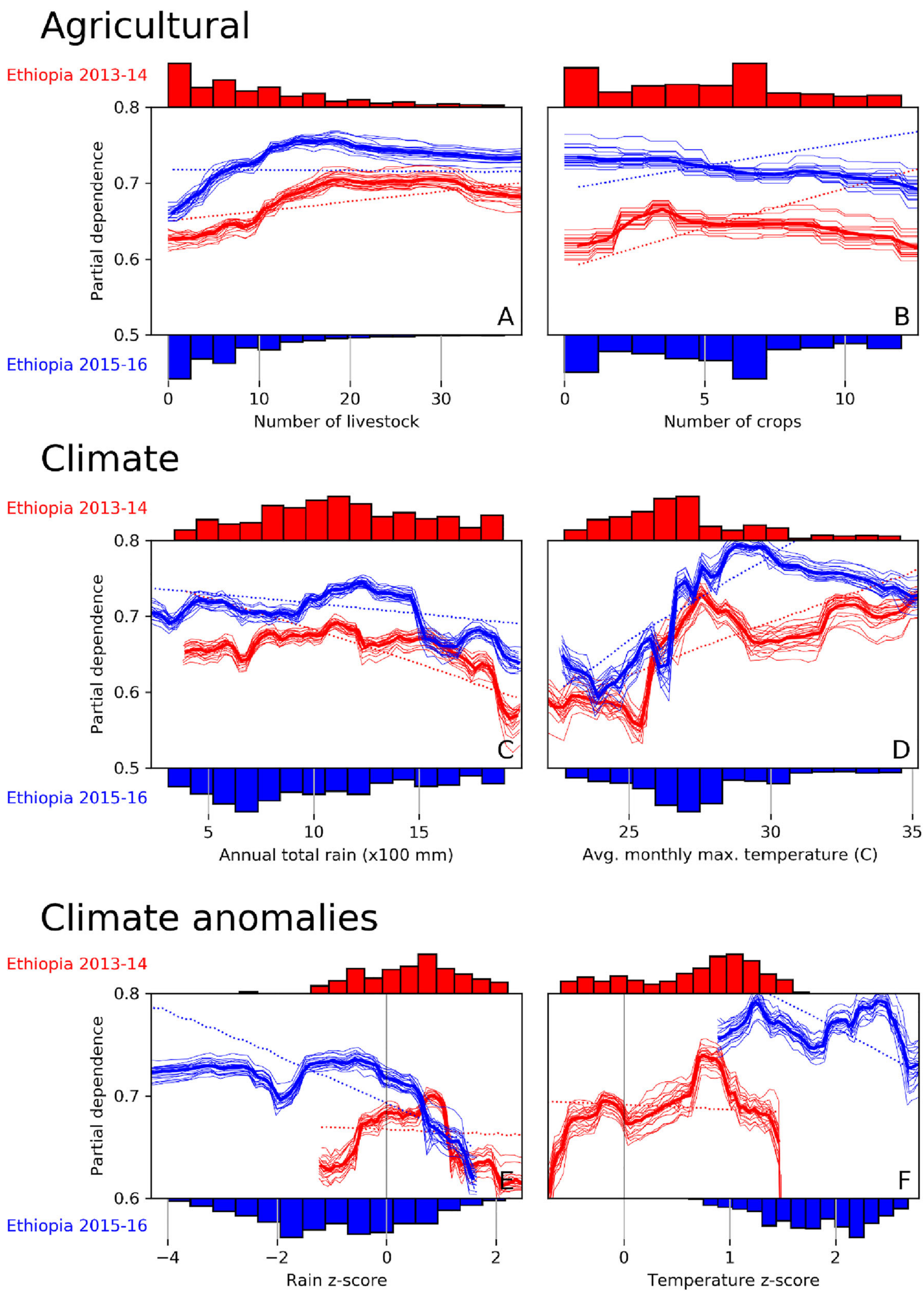
Livestock is an important household asset, and it has shown a positive relationship with various measures of food and nutritional security in previous research (Hoddinott et al. 2015; Shively and Sununtnasuk 2015; Slavchevska 2015). Here we examined for nonlinear patterns in that relationship. The results reveal a nonlinear relationship between the number of livestock and both measures of food security in Ethiopia in the 2014 and 2016 samples: owning up to about 18 large and small ruminants is associated with better self-reported food security and higher household dietary diversity, and after 18 head of livestock the relationship levels off (Figs. 1a and 2a). Since different kinds of livestock have different values for household food security (Sansoucy 1995), we also disaggregated livestock into the number of goats and sheep and the number of cows and oxen (Figure A-2 in the Supplementary Material). It appears that owning up to about 7 cows/oxen is linked to better self-reported food security, after which the positive relationship disappears. Although this threshold should be interpreted with caution because it is observed at the tail of the data distribution, it merits further investigation as it is in line with other work corroborating that livestock is associated with better food security (Porter 2012), but that there might be diminishing returns to livestock ownership (Dercon 2004). No clear thresholds are observed for the measures describing livestock ownership and food security outcomes in Nigeria (Figs. 3a and 4a).

4.2 Number of cultivated crops

In previous research, the number of cultivated crops has fairly consistently shown to be positively associated with dietary diversity, although the strength of the association is not always strong (Shively and Sununtnasuk 2015; Sibhatu et al. 2015; Slavchevska 2015). Our findings are consistent with these results, and we find no thresholds in that relationship for either country. A minor inflection point can be identified in 2014 for Ethiopia, when cultivating up to approximately 4 crops was associated with better self-reported food security (Fig. 1b). For Nigeria we observe that an association between the number of crops and self-reported food security status becomes negative after about 4 crops in 2013 (Fig. 3b). As a secondary analysis, we disaggregated the number of crops by the number of cereal and non-cereal crops and repeated the analyses. We found no major differences in the observed relationships between the number of cereal vs. non-cereal crops for either country (Figures A-2 and A-3 in the Supplementary Material).

4.3 Climatic conditions

Here we describe the associations between rainfall and temperature as well as temperature and rainfall anomalies



observed in the most recent agricultural season. In general, our results demonstrate several nonlinearities and inverted U-shaped relationships between the measures of climate and food security, suggesting there are optimal climatic conditions associated with positive food security outcomes.

◀ **Fig. 1** Partial dependence plots comparing the linear and non-parametric relationships between selected covariates and self-reported food security in Ethiopia based on the 2013–14 (red) and 2015–16 (blue) samples of the Ethiopia LSMS. The bars on top and bottom of the graphs represent histograms of sample distribution for a given variable for LSMS 2013–14 (top, red) and 2015–16 (bottom, blue). Higher partial dependence represents higher food security. The PDPs for the panel "Agricultural" and "Climate" were estimated using the Model 1 specification (described in main text). The PDPs for "Climate anomalies" were estimated using Model 2. All PDPs control for the full set of demographic and socioeconomic covariates described in the text but are not shown here to preserve space. 20 bootstrapped samples of the datasets were taken to generate the uncertainty bands for the random forest. For visual clarity, uncertainty bands in the linear model are not displayed and all data and PDPs were truncated at the 5% and 95% quantiles. Logistic regression coefficients and significance levels are represented in Tables A2-A3 in the Supplementary Material

In Ethiopia, there is a nonlinear association between total rainfall in the last 12 months and self-reported food security status in 2016 (Fig. 1c), with an inflection point of about 1200 mm (i.e., food security is highest at 1200 mm). In a similar way, the relationship between rainfall and self-reported food security in 2014 becomes negative after about 1500 mm of rainfall.

A non-linear relationship exists between rainfall and HDDS in 2016 in Ethiopia, with several inflection points: up to 1100 mm of rain in the preceding 12 months is associated with lower HDDS, but rainfall in the range of 1100–1800 mm is associated with more diverse diets for households as indicated by higher HDDS (Fig. 2c). This nonlinearity may be reflective of the drought conditions in the aftermath of which the Ethiopia 2015–16 LSMS was collected and potentially indicates that during that time, households consumed less diverse diets (lower HDDS). The relationship between rainfall and HDDS in 2014 becomes negative after ~1800 mm of rainfall (Fig. 2c).

Similar to the measure of rainfall in the last 12 months, we also observe nonlinear relationships between rainfall anomalies and self-reported food security in Ethiopia in both years (Fig. 1e), indicating that too little and too much rainfall relative to the long-term normal is negatively associated with self-reported household food security. In particular, the relationship between rainfall anomaly z-scores and self-reported food security resembles an inverted U-shape in 2014 (Fig. 1e); it looks nonlinear, although the exact shape is hard to ascertain, for self-reported food security and HDDS in 2016 (Figs. 1e and 2e).

In Nigeria, the relationship between total rainfall in the last 12 months and the two food security measures is noticeably more linear than for Ethiopia (Figs. 3c and 4c). Unlike in Ethiopia, the measure of recent rainfall shows a consistently positive relationship with HDDS (Fig. 4c). We also observe what looks like a weak inverted U-shaped relationship between rainfall anomalies z-scores and self-reported food security status for Nigeria in 2016 (Fig. 3e).

We observe inverted U-shape relationships between temperature anomalies and self-reported food security in Ethiopia in 2014 and 2016 (Fig. 1f). No clear thresholds are observed between the temperature anomalies and HDDS Ethiopia (Fig. 2f), although the direction of the relationship is opposite for the 2014 and 2016 survey rounds. There is no clear pattern that we can discern from the relationship between temperature z-score anomalies and self-reported food security in Nigeria (Fig. 3f), and the relationship between temperate anomalies and HDDS appears linear (Fig. 4f).

4.4 Limitations and extensions

The observed relationships are specific to the datasets that we used, and we do not claim that these represent trends generalizable to other data samples. PDPs implicitly assume that the covariate being varied is uncorrelated with all other covariates (i.e., that hypothetically varying it over its range would not affect the values of other covariates). When this assumption is violated, the resulting PDP is built from sets of values (**X**) that are very unlikely or even implausible in reality. Given that we do not observe large correlations between any of our covariates as indicated by the VIFs, we believe that this assumption is justified in our case.

The numerical values of thresholds (e.g., after 18 livestock the relationship levels off) are approximate and should not be interpreted in a prescriptive sense. The goal of this study was to identify if such thresholds/nonlinearities exist, but in order to establish a precise threshold value, a different research design – a randomized experiment, for example – is needed.

We utilized two measures of food security for our analysis. The self-reported measure is advantageous as it is simple and dichotomous, whereas HDDS is also easy to understand and collect. Yet these measures do not fully represent the multidimensional issue of food (in)security, and there exist a variety of different methods for quantifying food security, each of which has advantages and disadvantages (Headey and Ecker 2013; Jones et al. 2013). For example, we observe flat curves for the associations between the number of cultivated crops and HDDS (they appear even flatter than those for the self-reported measure of food security). However, this might be a feature of the HDDS as a measure – an additional cereal crop may represent an additional food source, but if cereals are already present in a household diet, it does not represent a new food group that would increase the HDDS for a given household. Even though we did not find differences in the associations when we disaggregated the number of cultivated crops to the number of cereal and non-cereal crops, more research is needed to better understand the relationships between the types of cultivated crops and household food security. Relatedly, addressing multiple dimensions

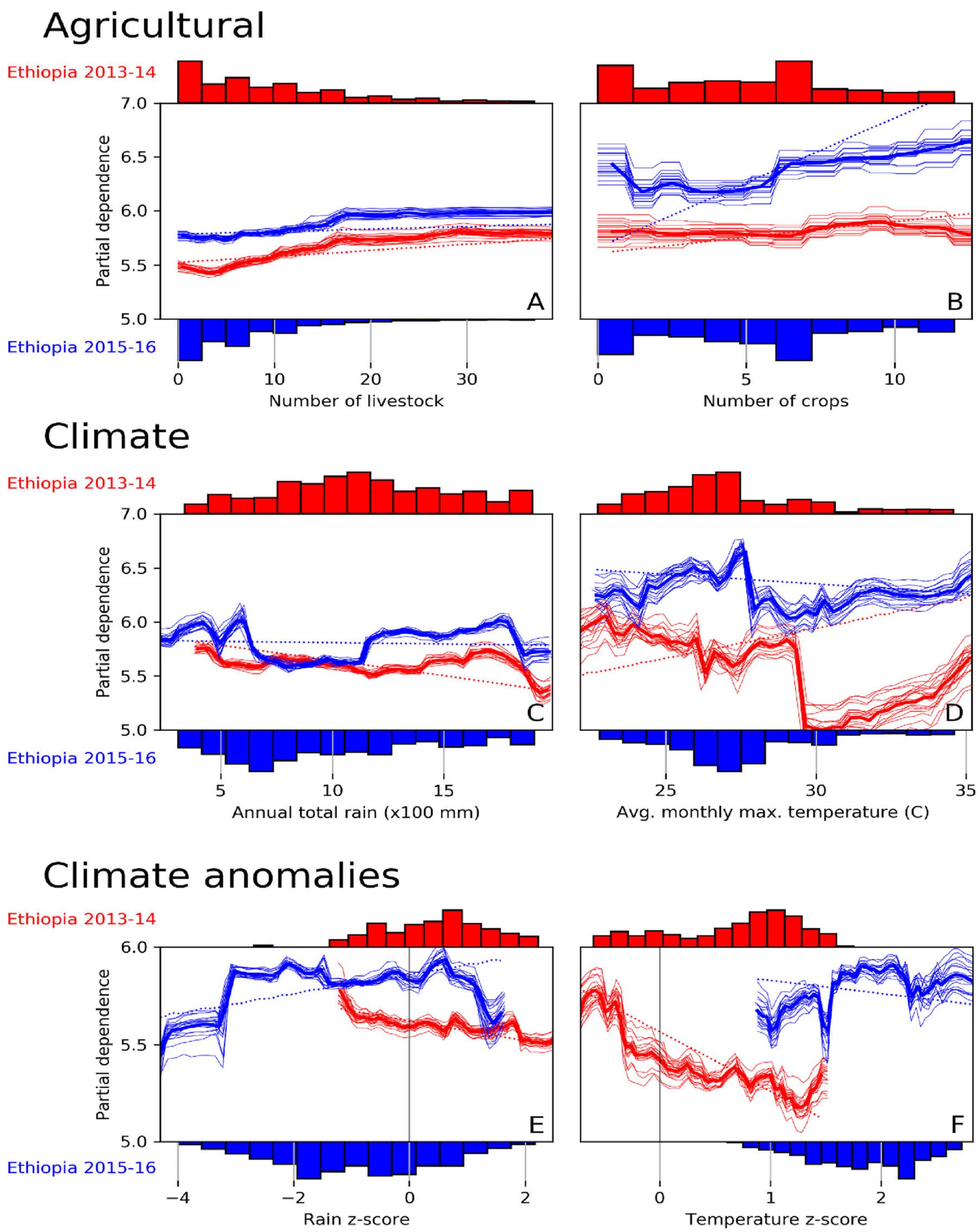


Fig. 2 Partial dependence plots comparing the linear and non-parametric relationships between selected covariates and household dietary diversity score (HDDS) in Ethiopia based on the 2013–14 (red) and 2015–16 (blue) samples of the Ethiopia LSMS. Linear regression coefficients and significance levels are represented in Tables A2-A3. Please refer to Fig. 1 for technical details about the figure

of food security by utilizing its various definitions and measures within the nonparametric framework is a promising direction for future research.

The suite of nonparametric methods that we explored is by no means comprehensive. Prior to selecting the random forest (see Supplemental Material), we aimed to demonstrate several commonly used methods that are simple to implement in R (and other languages). It is possible that substantial improvements in predictive accuracy could be gained by exploring more methods, ensembles of the methods used, or through hyperparameter selection. Future work could aim to explore these avenues.

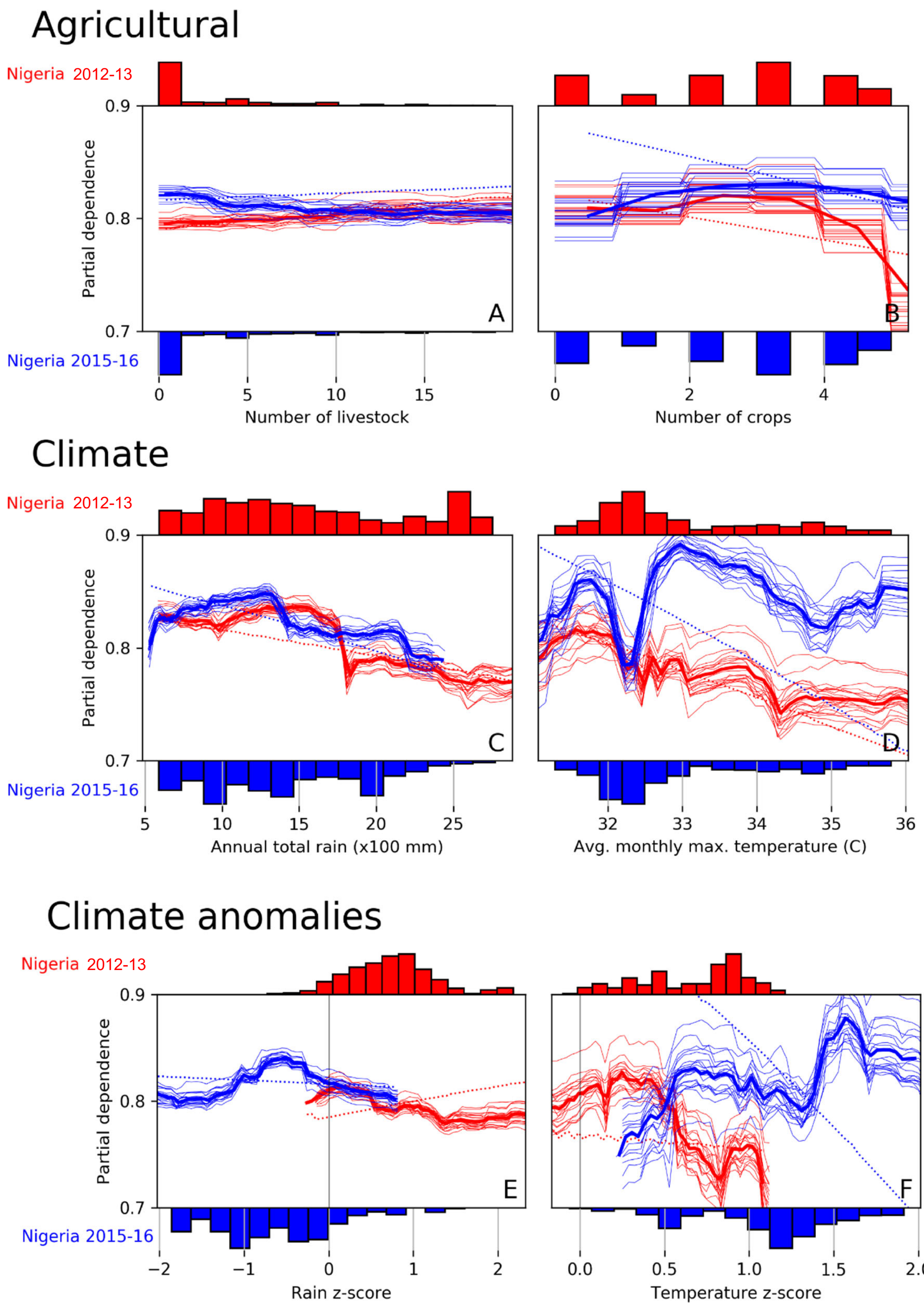


Fig. 3 Partial dependence plots comparing the linear and non-parametric relationships between selected covariates and self-reported food security in Nigeria based on the 2012–13 (red) and 2015–16 (blue) samples of the Nigeria LSMS. Logistic regression coefficients and significance levels are represented in Tables A4-A5. Please refer to Fig. 1 for technical details about the figure

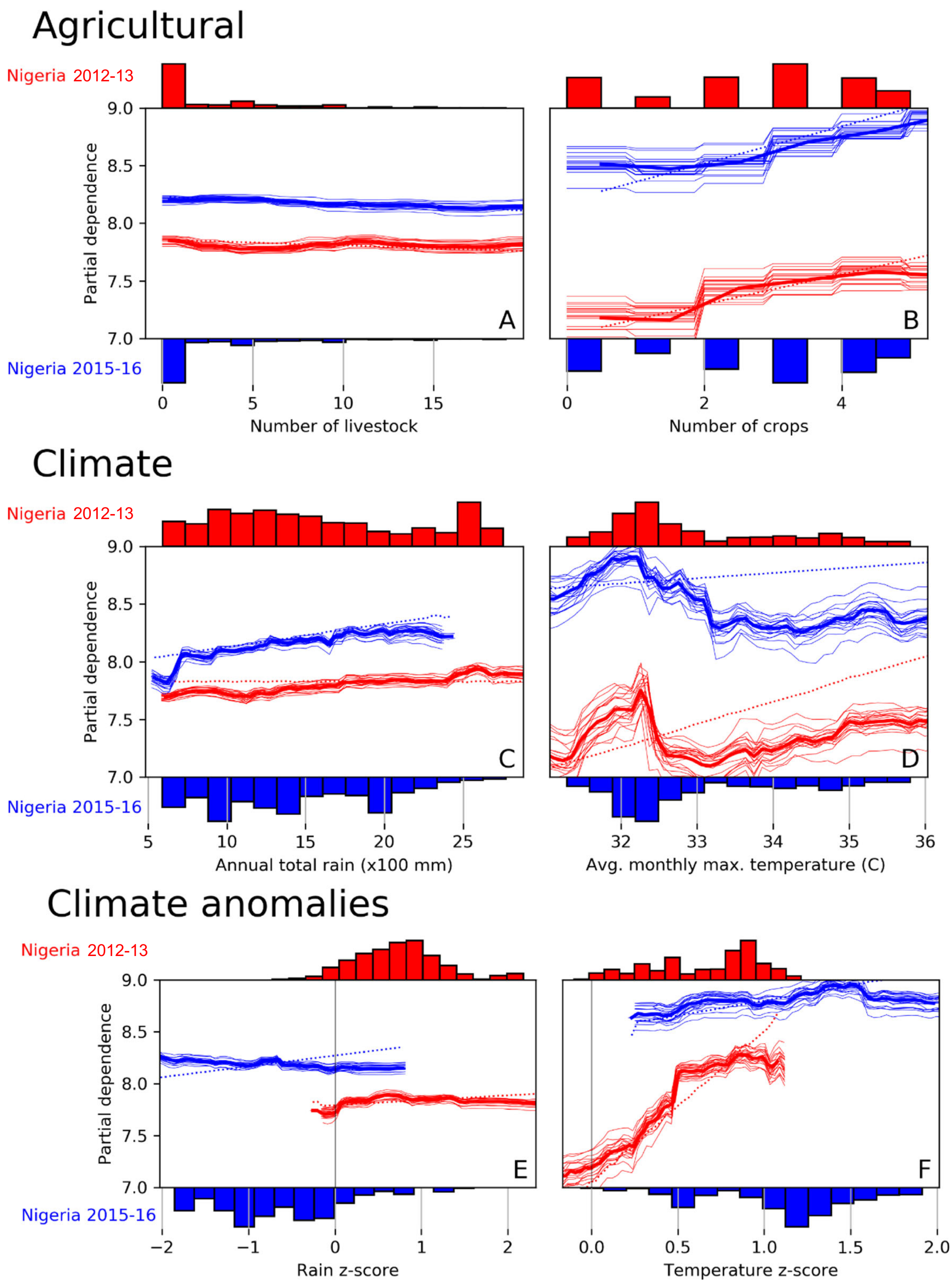


Fig. 4 Partial dependence plots comparing the linear and non-parametric relationships between selected covariates and household dietary diversity score (HDDS) in Nigeria in 2012–13 (red) and 2015–16 (blue) of the Nigeria LSMS. Linear regression coefficients and significance levels are represented in Tables A4-A5. Please refer to Fig. 1 for technical details about the figure

5 Discussion

The goal of this study was to analyze and describe the relationships between self-reported food security status, household dietary diversity score, and their determinants by using traditional methods—linear and logistic regressions—and a nonparametric method—random forest. As the results show, random forest can help discover nuances such as nonlinearities and thresholds, which could further be used to identify vulnerable populations and inform theory about the determinants of food security. Several substantive and policy-relevant findings and their implications should be mentioned. First, a general positive association with a threshold of about 18 (small and large ruminants) is observed between livestock ownership and both self-reported food security status and household dietary diversity score in Ethiopia. While the positive relationship between livestock ownership and food security is not a novel finding, the fact that we observe this association in years with normal rainfall and in the aftermath of a drought may indicate that livestock ownership, particularly the number of cows and oxen, can be protective of household food security during droughts. Further, this threshold may point to diminishing returns of livestock ownership for household food security, which has been observed in previous research (Dercon 2004). Given that the LSMS survey does not sample a sufficiently large population of pastoralists (those primarily engaged in animal husbandry), our findings regarding livestock in both countries are likely specific to mixed cropping-livestock smallholder farming systems, warranting further research into potential nonlinear relationships between herd size and food security in pastoralist communities. Second, a relatively flat relationship between the number of cultivated crops (as opposed to the non-linear relationship with livestock) and the measures of food security indicates that crops potentially contribute to household food security in a fundamentally different way than livestock. While we can only speculate why this might be the case, this finding warrants further research into the pathways connecting household agricultural livelihoods and food security. Taken together, these findings indicate that policies aimed at increasing agricultural productivity may be beneficial for household food security outcomes only up to a point.

Our study also demonstrates nonlinear relationships between climatic conditions and food security outcomes in Ethiopia, while in Nigeria the relationships are more linear. In particular, we show that too much and too little rainfall is negatively correlated with the self-reported measure of household food security in both countries. In addition, higher-than-average temperature (temperature anomalies) in Ethiopia are strongly negatively associated with self-reported food security and HDDS in both survey samples. Non-linearity in the relationship between rainfall and self-reported food security and HDDS in the 2016 sample in Ethiopia may in fact reflect the observed drought conditions in Ethiopia that prevailed before the survey. As such, the random forest may be useful for uncovering and illustrating the patterns among the determinants of food security based on surveys conducted in different years and in different climatic conditions. While several recent studies showed no large-scale quantitative impact of the 2015 drought in Ethiopia on agricultural production, wages, prices, and the rates of child malnutrition (Hirvonen et al. 2020; Sohnesen 2019), our findings could be used to inform future research on the response of rural households to droughts.

In summary, we show that a random forest can be used to enhance insight provided by the traditional methods in food security research and advocate for a wider use and implementation of nonparametric methods into food security analyses. Our study provides insight about thresholds among various measures of household agricultural activities, climatic conditions, and food security, which, at the very least, warrants future research into the complex issue of household food security. The threshold-based relationships observed in our analysis suggest that policies aimed at increasing agricultural productivity (e.g., livestock holdings) may only be beneficial up to an extent. While it is intuitive that some level of diminishing returns will exist, our analysis could be used as a first step to discern the levels to which policies may be beneficial. More generally, our results reveal instances in which linear analysis does not capture nuances in the relationships in the data. Our findings show the kind of analysis that we have presented in this work could complement traditional forms of statistical inference in social science research and beyond.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12571-020-01132-w.

Acknowledgments The authors express gratitude to the World Bank, Nigeria Bureau of Statistics, Ethiopia Central Statistical Agency, and the Climate Hazard Center at UC Santa Barbara for making the data used in this study publicly available. We thank Kathryn Grace for her helpful comments.

Authors' contributions MB, TGW, and SDG designed the study; MB, TGW, and AV processed data and performed statistical analyses; MB, TGW, SV, and wrote the paper.

Funding The authors gratefully acknowledge support from a National Science Foundation INFEWS grant (Award #1639214). MB and AV are grateful for the support from the Minnesota Population Center (#R24 HD041023), funded through grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. MB has done most of the work on the manuscript while she was a PhD student at the University of Minnesota.

Data availability Survey data are publicly available at https://microdata.worldbank.org/index.php/catalog/lsms#_r=&collection=&country=66&dtype=&from=1890&page=1&ps=&sid=&sk=&sort_by=nation&sort_order=&to=2017&topic=&view=s&vk=

Climate data are publicly available at https://www.chc.ucsb.edu/data

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Code availability (software application or custom code) R code to perform the analysis is available upon request.

References

Abafita, J., & Kim, K.-R. (2014). Determinants of household food security in rural Ethiopia: an empirical analysis. *Journal of Rural Development/Nongchon-Gyeongje, 37*(1071-2016-86950), 129–157.

Abay, K., & Hirvonen, K. (2017). Does market access mitigate the impact of seasonality on child growth? Panel data evidence from northern Ethiopia. *The Journal of Development Studies, 53*(9), 1414–1429.

Babatunde, R. O., & Qaim, M. (2010). Impact of off-farm income on food security and nutrition in Nigeria. *Food Policy, 35*(4), 303–311.

Barrett, C. B., & Dorosh, P. A. (1996). Farmers' welfare and changing food prices: nonparametric evidence from rice in Madagascar. *American Journal of Agricultural Economics, 78*(3), 656–669.

Beyene, F., & Muche, M. (2010). Determinants of food security among rural households of Central Ethiopia: an empirical analysis. *Quarterly Journal of International Agriculture, 49*(4), 299.

Bogale, A., & Shimelis, A. (2009). Household level determinants of food insecurity in rural areas of Dire Dawa, eastern Ethiopia. *African Journal of Food, Agriculture, Nutrition and Development, 9*(9).

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Carletto, G., Ruel, M., Winters, P., & Zezza, A. (2015). Farm-level pathways to improved nutritional status: introduction to the special issue. *The Journal of Development Studies, 51*(8), 945–957. https://doi.org/10.1080/00220388.2015.1018908.

Deaton, A. (1989). Rice prices and income distribution in Thailand: a non-parametric analysis. *The Economic Journal, 99*(395), 1–37.

Demeke, M., Guta, F., Ferede, T., & Ababa, A. (2004). Agricultural development in Ethiopia: Are there alternatives to food aid? *Department of Economics, Addis Ababa University*. www.sarpn.org.za/documents/d0001583/FAO2005_Casestudies_Ethiopia.pdf

Demeke, A. B., Keil, A., & Zeller, M. (2011). Using panel data to estimate the effect of rainfall shocks on smallholders food security and vulnerability in rural Ethiopia. *Climatic Change, 108*(1–2), 185–206.

Deng, H., Runger, G., & Tuv, E. (2011). Bias of importance measures for multi-valued attributes and solutions. In *International conference on artificial neural networks* (pp. 293–300). Berlin, Heidelberg: Springer.

Dercon, S. (2004). Growth and shocks: evidence from rural Ethiopia. *Journal of Development Economics, 74*(2), 309–329.

Devendra, C. (2005). *Small ruminants in Asia; Contribution to food security, poverty alleviation and opportunities for productivity enhancement.* Paper presented at the Proceeding of international workshop on small ruminant production and development in South East Asia. MEKARN, Nong Lam, HCMC, Vietnam.

Devereux, S., & Sussex, I. (2000). *Food insecurity in Ethiopia*. Institute for Development Studies.

EPAR. (2019). AgDev priority indicators: Summary statistics round 2 (v7.2). In Seattle: University of Washington, Evans School Policy Analysis and Research Group (EPAR).

FAO. (1996). *Rome declaration on world food security and world food summit plan of action*. Rome: FAO.

Feleke, S. T., Kilmer, R. L., & Gladwin, C. H. (2005). Determinants of food security in southern Ethiopia at the household level. *Agricultural Economics, 33*(3), 351–363.

Frelat, R., Lopez-Ridaura, S., Giller, K. E., Herrero, M., Douxchamps, S., Djurfeldt, A. A., et al. (2016). Drivers of household food availability in sub-Saharan Africa based on big data from small farms. *Proceedings of the National Academy of Sciences, 113*(2), 458–463.

Funk, C., Rowland, J., Eilerts, G., Kebebe, E., Biru, N., White, L., & Galu, G. (2012). A climate trend analysis of Ethiopia. *US Geological Survey, Fact Sheet, 3053*.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data, 2*(1), 150066. https://doi.org/10.1038/sdata.2015.66.

Funk, C., Peterson, P., Peterson, S., Shukla, S., Davenport, F., Michaelsen, J., et al. (2019). A high-resolution 1983–2016 T max climate data record based on infrared temperatures and stations by the climate Hazard center. *Journal of Climate, 32*(17), 5639–5658.

Grace, K., Husak, G., Harrison, L., Pedreros, D., & Michaelsen, J. (2012). Using high resolution satellite imagery to estimate cropped area in Guatemala and Haiti. *Applied Geography, 32*(2), 433–440.

Grace, K., Husak, G., & Bogle, S. (2014). Estimating agricultural production in marginal and food insecure areas in Kenya using very high resolution remotely sensed imagery. *Applied Geography, 55*, 257–265. https://doi.org/10.1016/j.apgeog.2014.08.014.

Headey, D., & Ecker, O. (2013). Rethinking the measurement of food security: from first principles to best practice. *Food Security, 5*(3), 327–343.

Headey, D. D., & Hoddinott, J. (2016). Agriculture, nutrition and the green revolution in Bangladesh. *Agricultural Systems, 149*, 122–131.

Hirvonen, K., Sohnesen, T. P., & Bundervoet, T. (2020). Impact of Ethiopia's 2015 drought on child undernutrition. *World Development, 131*, 104964.

Hoddinott, J., Headey, D., & Dereje, M. (2015). Cows, missing milk markets, and nutrition in rural Ethiopia. *The Journal of Development Studies, 51*(8), 958–975.

Husak, G., & Grace, K. (2016). In search of a global model of cultivation: using remote sensing to examine the characteristics and constraints of agricultural production in the developing world. *Food Security, 8*(1), 167–177. https://doi.org/10.1007/s12571-015-0538-6.

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, 353*(6301), 790–794.

Jones, A. D. (2017). Critical review of the emerging research evidence on agricultural biodiversity, diet diversity, and nutritional status in low- and middle-income countries. *Nutrition Reviews, 75*(10), 769–782.

Jones, A. D., Ngure, F. M., Pelto, G., & Young, S. L. (2013). What are we assessing when we measure food security? A compendium and review of current metrics. *Advances in Nutrition, 4*(5), 481–505.

Kugler, T. A., Grace, K., Wrathall, D. J., de Sherbinin, A., Van Riper, D., Aubrecht, C., et al. (2019). People and pixels 20 years later: the current data landscape and research trends blending population and environmental data. *Population and Environment, 41*(2), 209–234.

Kuku, O., Garasky, S., & Gundersen, C. (2011). The relationship between childhood obesity and food insecurity: a nonparametric analysis. *Applied Economics, 44*(21), 2667–2677.

Lentz, E., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development, 122*, 399–409.

Lobell, D. B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., & Murray, S. (2020). Eyes in the sky, boots on the ground: assessing satellite-and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics, 102*(1), 202–219.

Müller, B., Hoffmann, F., Heckelei, T., Müller, C., Hertel, T. W., Polhill, J. G., et al. (2020). Modelling food security: bridging the gap between the micro and the macro scale. *Global Environmental Change, 63*, 102085.

Myers, S. S., Smith, M. R., Guth, S., Golden, C. D., Vaitla, B., Mueller, N. D., Dangour, A. D., & Huybers, P. (2017). Climate change and global food systems: potential impacts on food security and undernutrition. *Annual Review of Public Health, 38*, 259–277.

Nawrotzki, R., DeWaard, J., Bakhtsiyarava, M., & Ha, J. T. (2017). Climate shocks and rural-urban migration in Mexico: exploring nonlinearities and thresholds. *Climatic Change, 140*(2), 243–258. https://doi.org/10.1007/s10584-016-1849-0.

Philip, S., Kew, S. F., Jan van Oldenborgh, G., Otto, F., O'Keefe, S., Haustein, K., et al. (2018). Attribution analysis of the Ethiopian drought of 2015. *Journal of Climate, 31*(6), 2465–2486.

Porter, C. (2012). Shocks, consumption and income diversification in rural Ethiopia. *Journal of Development Studies, 48*(9), 1209–1222.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. URL https://www.R-project.org/

Reardon, T. (1997). Using evidence of household income diversification to inform study of the rural nonfarm labor market in Africa. *World Development, 25*(5), 735–747.

Sam, A. S., Abbas, A., Padmaja, S. S., Kaechele, H., Kumar, R., & Müller, K. (2019). Linking food security with household's adaptive capacity and drought risk: implications for sustainable rural development. *Social Indicators Research, 142*(1), 363–385.

Sansoucy, R. (1995). Livestock-a driving force for food security and sustainable development. *World, 3074*(5389), 1035.

Schlenker, W., & Lobell, D. B. (2010). Robust negative impacts of climate change on African agriculture. *Environmental Research Letters, 5*(1), 014010.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of Sciences, 106*(37), 15594–15598.

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*(50), 19703–19708. https://doi.org/10.1073/pnas.0701976104.

Shively, G., & Sununtnasuk, C. (2015). Agricultural diversity and child stunting in Nepal. *Journal of Development Studies, 51*(8), 1078–1096. https://doi.org/10.1080/00220388.2015.1018900.

Shortridge, J. E., Falconi, S. M., Zaitchik, B. F., & Guikema, S. D. (2015). Climate, agriculture, and hunger: statistical prediction of undernourishment using nonlinear regression and data-mining techniques. *Journal of Applied Statistics, 42*(11), 2367–2390.

Sibhatu, K. T., & Qaim, M. (2017). Rural food security, subsistence agriculture, and seasonality. *PLoS One, 12*(10), e0186406.

Sibhatu, K. T., Krishna, V. V., & Qaim, M. (2015). Production diversity and dietary diversity in smallholder farm households. *Proceedings of the National Academy of Sciences, 112*(34), 10657–10662.

Slavchevska, V. (2015). Agricultural production and the nutritional status of family members in Tanzania. *The Journal of Development Studies, 51*(8), 1016–1033.

Sohnesen, T. P. (2019). Two sides to same drought: measurement and impact of Ethiopia's 2015 historical drought. *Economics of Disasters and Climate Change*, 1–19.

Swindale, A., & Bilinsky, P. (2006). *Household dietary diversity score (HDDS) for measurement of household food access: Indicator guide*. Washington, DC: Food and Nutrition Technical Assistance Project, Academy for Educational Development.

UNDP. (2015). Sustainable development goals. United Nations Development Programme.

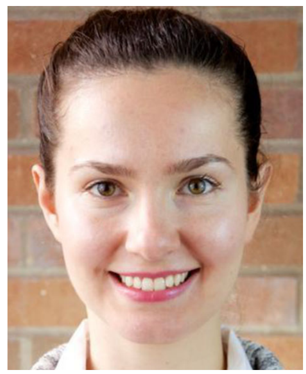
Webb, P., & Kennedy, E. (2014). Impacts of agriculture on nutrition: nature of the evidence and research gaps. *Food and Nutrition Bulletin, 35*(1), 126–132. https://doi.org/10.1177/156482651403500113.

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science, 341*(6145), 508–513.

Workneh, A., Getahun, E., Tibbo, M., Mamo, Y., & Rege, J. (2003). *Current state of knowledge on characterization of farm animal genetic resources in Ethiopia. Farm animal biodiversity in Ethiopia: Status and prospects.* Paper presented at the Asfaw Yimegnuhal and Tamrat Degefa (eds), Proceedings of the 11th Annual Conference of the Ethiopian Society of Animal Production (ESAP) held in Addis Ababa, Ethiopia, 28–30 August 2003.

World Bank. (2019). LSMS-integrated surveys on agriculture. Available at: http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA: World Bank.

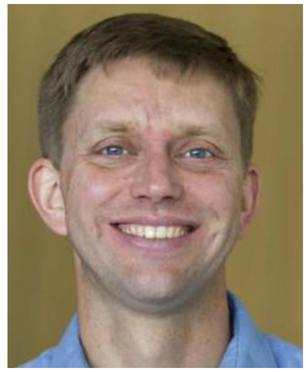
Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., et al. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications, 11*(1), 1–11.



Maryia Bakhtsiyarava is a postdoctoral scholar at the Institute of Urban and Regional Development at the University of California – Berkeley. She completed her PhD at the Department of Geography, Environment and Society at the University of Minnesota in Summer 2020. Maryia's doctoral work analyzed how household-level food (in)security interacts with children's nutrition outcomes and relates to households' livelihood strategies manifested through their agricultural activities in the context of climate variability.

Tim G. Williams is a PhD Candidate at the Department of Industrial and Operations Engineering at the University of Michigan. Tim's underlying motivation is to encourage agricultural systems to develop and transform in a way that promotes both the betterment of society and the prosperity of the environment. Tim uses both agent-based modeling to examine how systems may behave in the future or under intervention and empirical analysis and data mining to better understand observed historical relationships.

Dr. Andrew Verdin participated in this research as a research scientist at the Minnesota Population Center. Andrew's current research interests include: (i) developing innovative methods for quantifying the effect of climate extremes on maternal and child health, (ii) providing guidance to properly merge climate and population data, (iii) assessing subsistence farming sustainability in rural Mali, and (iv) analyzing the effects of environment on circular migration in rural Mali.



Dr Guikema's research is focused on and grounded in risk analysis, statistical learning theory, agent-based modeling, stochastic simulation, decision analysis, and agent-based modeling. Dr. Guikema and his research group focus on issues related to climate adaptation and the sustainability of cities and infrastructure, disaster risk analysis, critical infrastructure modeling, natural hazards, and terrorism risk. His work is a mix of developing and testing new methods and applying methods to better understand problems of societal importance.