Working paper MOZ-22195

August 2024

Predicting the impact of global warming on vulnerability to food insecurity in rural Mozambique

Paulo Santos
Reiss McLeod

IGC

DIRECTED BY LSE UNIVERSITY OF OXFORD

FUNDED BY UKaid from the British people

Predicting the impact of global warming on vulnerability to food insecurity in rural Mozambique

Paulo Santos and Reiss McLeod

Monash University

Abstract

We combine household data from rural areas of Mozambique with detailed information on the local physical environmental to characterize the relation between food security and the environment in rural areas of Mozambique. Using machine learning techniques (forest trees and associated surrogate models), we characterize the heterogeneity in the profiles of vulnerability to food insecurity as well as its environmental best predictors. Temperature in the first months of the main production season is the main predictor of food consumption score, but the effect of this variable is nonlinear. We use the predictive nature of the surrogate model to quantify the prevalence of food insecurity under different temperature scenarios. Our results highlight the central role of livelihood transitions (and the associated role of policy) in determining the outcomes of global warming.

1 Introduction

Mozambique is considered highly vulnerable to climate change given its geographical location, its long coastline and the importance of the area close to (or below) mean sea level. Global warming may impact the welfare of rural populations through rising sea levels, as well as the increase in the intensity and frequency of natural hazards, including droughts, floods and cyclones. While these shocks have deservedly received much attention, changes in average temperature may also lead to other, slow moving, impacts, via changes in the profitability and risk of current agricultural production. Because these changes have received less attention, particularly when relying on detailed microeconomic data, their effects are less well understood. This report is one first effort to address this gap.

In the next section we briefly present the household data used in our analysis, the Inquerito Agricola Integrado 2012-2014 (IAI2012), as well as the indicator of food security used in our analysis, the Food Consumption Score. In addition to its statistical design, that makes it representative of rural areas of Mozambique, our data includes precise information on household's location, allowing us to link the socio-economic data with a wide array of other datasets of spatially explicit environmental variables, including climatic variables. These data are also presented in section 2.

Section 3 presents our empirical approach. Following a well-established literature in both production economics and poverty measurement, we use the characterization of the conditional distribution of food security (ie, estimates of the conditional mean and variance of the food consumption score), and characterize vulnerability as the empirical probability of a household with specific characteristics to have a diet classified as below acceptable. We hypothesise that the relation between food security and environmental characteristics is essentially heterogeneous and use machine learning techniques to explore this heterogeneity.

Our results, presented in section 4, show that households in rural areas of Mozambique can be classified into eleven more homogeneous groups, with significantly different levels of food security. These groups reflect environmental production conditions (and, implicitly, households' choice of livelihood strategies, including agricultural production, as an adaptation to those conditions). Importantly, this result suggests a nonlinear relation between food security and average temperature at critical times of the main agricultural season. While households living in cooler areas (characterised as those with an average temperature during the first two months of the main agricultural season below 25C) are, on average, more food secure than those living in warmer areas, there is an important exception: households living in very hot areas (average temperature during the first two months of the main agricultural season *above* 29C) are exceptionally well-off, possibly reflecting their apparent specialisation in intensive cattle production.

We then use climate scenarios to predict future (2050, 2070 and 2090) vulnerability to food insecurity, assuming different alternative livelihood options (ie, when we include/exclude the possibility of adapting to increases in temperature by practicing the set of activities practiced by the cluster of households in the warmest cluster identified in the data). These simulations illustrate that what is considered as feasible adaptation strategies can have a large effect on the predictions of the effect of global warming, even in the absence of technological change. While under a restricted set of options, global warming leads to a reduction in food security, the opposite is true when we assume that rural households can adapt by adopting, without restrictions, current technologies. We discuss the limitations of this analysis and further work suggested by this analysis in the final section of this report.

2 Data and methods

2.1 Data

We use data from the Inquerito Agrario Integrado 2012 PME (IAI2012), a nationally representative survey of 6708 households distributed across 11 provinces and 146 districts, to quantify food insecurity in rural Mozambique. In addition to information on household demographic characteristics and assets, and detailed information on agricultural production, the IAI2012 included data on food consumption which we will use to calculate the Food Consumption Score, our indicator of food insecurity.

Importantly, this survey includes detailed information about the geographic location of each household. This allows us to to create a spatially explicit dataset on food insecurity and climatic variables for rural Mozambique in 2012, by linking the household data from the IAI2012 to spatial data on a large set of environmental variables that are potentially important predictors of agricultural production (and consumption) and, in the case of climatic variables, are predicted to change in the future.

2.1.1 Food insecurity

We measure dietary adequacy of food consumption by estimating the food consumption score (FCS). The FCS represents the frequency of food consumption across different food groups, including starches, pulses, vegetables, fruit, meat, dairy, fats and sugar in the 7 days prior to the survey (World Food Programme, 2024). Respondents are asked to recall the number of days they consumed different foods, from a list of 14 food items, with each item in the list corresponding to a food group. The crosswalk between items and food group is used to identify the number of days a food was eaten from each food group, with the maximum value for any food group capped at 7. The number of days each food group (Staples - St, Pulses - P, Vegetables - V, Fruits - Fr, Meat - M, Dairy – D, Fats – Fa, Sugars - Su) is consumed is weighted by their nutritional content leading to the food consumption score. We calculate the FCS using the expression:

$$FCS = 2St + 3P + V + Fr + 4(M + D) + .5(Fa + Su) \qquad (1)$$

The FCS can be used as a continuous indicator or used to define categories of diet quality: we define a household as being food insecure if their FCS is less or equal to 35, the threshold at above which a diet can considered adequate. Figure *1* presents the distribution of the FCS in our sample: worryingly, and by this measure, approximately 55 percent of the households are food insecure.

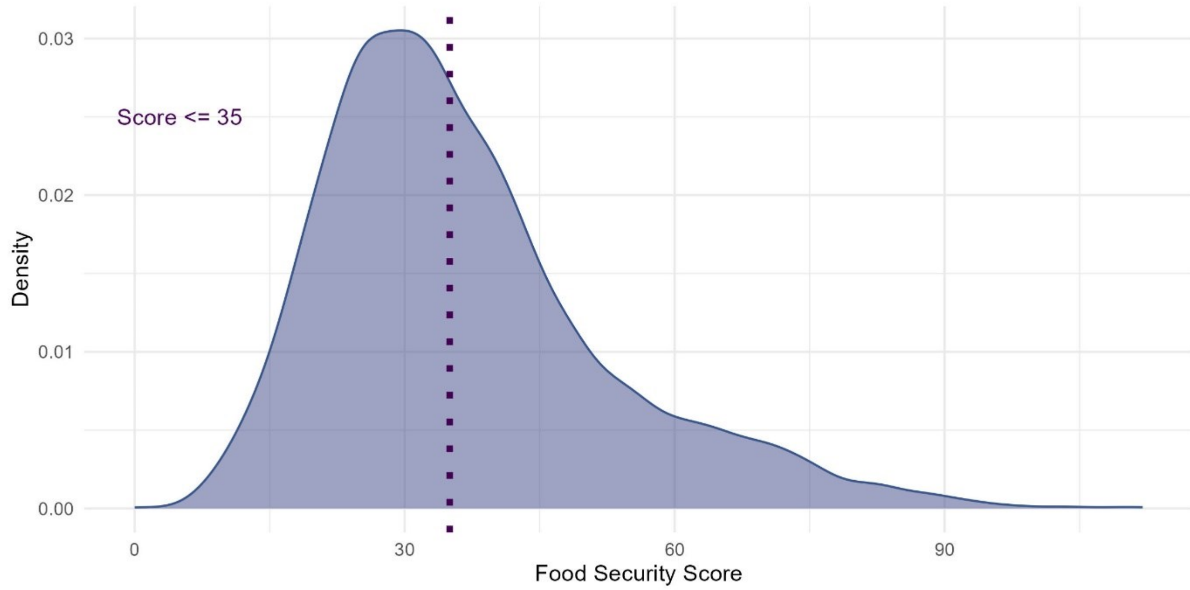


Figure 1 Distribution of the Food Consumption Score in IAI2012.

2.1.2 Spatial data

We characterize the environmental conditions affecting households surveyed in IAI2012 using different spatial datasets with global coverage. We start by creating buffer zones with a radius of 5000 metres centred on household locations, as provided by IAI2012 (see Figure 2) and calculate the mean of each variable within this buffer area for each spatial dataset. In most cases this process is performed using Google Earth Engine (GEE), and the results are exported for further analysis in R (see methods section). When data is missing for a household, we replace the missing data. with the sample mean.

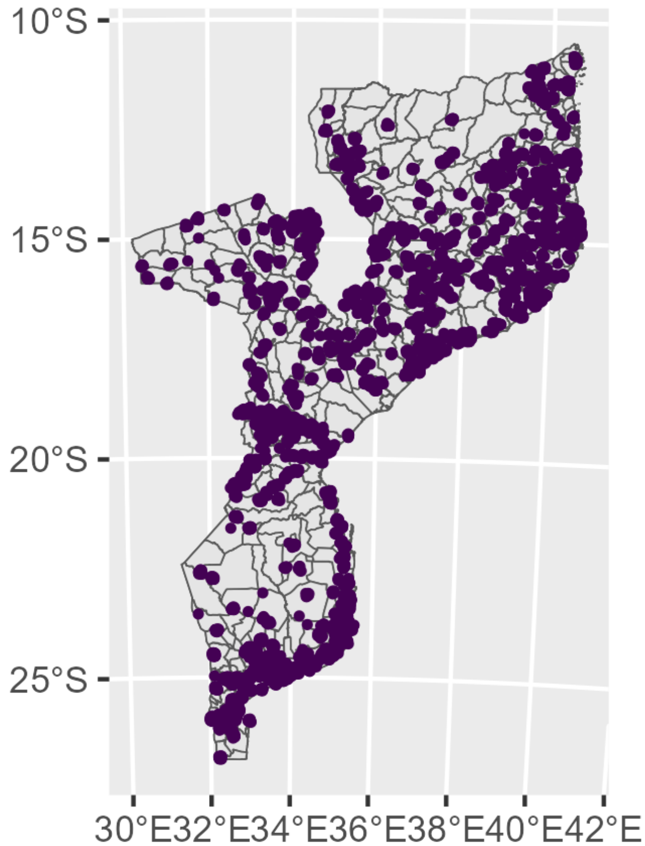


Figure 2 Household location (Note: these are approximate locations; exact locations were masked)

2.1.2.1 Natural capital

We characterize household's natural capital as the set of physical characteristics of the terrain, including elevation, slope and ruggedness, and soil properties (such as soil ph and soil carbon). Many of these variables, described in Table 1, are available on GEE.

Table 1 Natural capital variables

Variable	Description	Resolution	Year
Elevation	MERIT DEM is a high accuracy global DEM at 3 arc second resolution (~90 m at the equator) produced by eliminating major error components from existing DEMs (NASA SRTM3 DEM, JAXA AW3D DEM, Viewfinder Panoramas DEM) (Yamazaki *et al.*, 2017).	90m	2017
Slope	We calculate slope in degrees from MERIT DEM. The local gradient is computed using the 4-connected neighbors of each pixel. The slope at the selected pixel is calculated based on the elevation differences between the selected pixel and each of its four neighbors.	90m	2017

Variable	Description	Resolution	Year
Ruggedness	We use the high-resolution global Terrain Ruggedness Index data compiled by Nunn and Puga (2012) following the approach suggested by Riley et al. (1999) and calculated with data from GTOPO30 (USGS 1996) elevation data. GTOPO30 is a global elevation data set developed through a collaborative international effort led by staff at the US Geological Survey's Center for Earth Resources Observation and Science (EROS). The Terrain Ruggedness Index (TRI) is calculated as: $TRI_{r,c} = \sqrt{\sum_{i=r-1}^{r+1} \sum_{i=c-1}^{c+1} (e_{i,j} - e_{r,c})^2}$ where $e_{r,c}$ is the elevation in row r and cell c of the global elevation matrix of cells. TRI is the square root of the sum of all squared differences of the elevation of a grid from the elevation of its 8 surrounding grids.	30 Arc Seconds	1996
Soil carbon	This indicator provides a measure of the soil organic carbon content in soils at a depth of 30cm. The unit of measure is tonnes of carbon per hectare of soil, in the top 30 cm of soil. The resolution of the data is 250 metres (Hengl and Ichsani Wheeler, 2018).		
Soil ph	This indicator provides a measure of the soil ph in H2O at a depth of 30cm. The resolution of the data is 250 metres. The scale ranges from 0 to 140 (Hengl, 2018).		

2.1.2.2 Climate

We create historical climatic variables using ERA5 (**E**uropean Centre for Medium-Range Weather Forecasts **R**eanalysis **A**tmosphere 5th generation) daily data (Copernicus Climate Change Service, 2023) and create future climatic variables using Coupled Model Intercomparison Project climate projection data (Copernicus Climate Change Service, 2021). These historical climatic variables are processed using GEE, while future climate variables are processed using R.

ERA5 is the fifth generation ECMWF atmospheric reanalysis of the global climate. Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset. ERA5 DAILY provides aggregated values for each day for seven ERA5 climate reanalysis parameters: 2m air temperature, 2m dewpoint temperature, total precipitation, mean sea level pressure, surface pressure, 10m u-component of wind and 10m v-component of wind, and the data is available from 1979 to three months from real-time. We use data on 2m air temperature (presented as daily averages) and total precipitation (presented as daily sums) in our analysis.

CMIP climate projection data provides daily projections on several climatic variables for 2015 to 2100. We use data for SSP1-2.6, SSP2-4.5 and SSP5-8.5, where SSP1, SSP 2 and SSP5 refer to shared socioeconomic pathways 1, 2, and 5 respectively, and Representative Concentration Pathways (RCPs) 2.6, 4.5 and 8.5 refer to levels of radiative forcing in watts per square metre (W/m^2) by 2100. Generally, positive radiative forcing leads to warming of the Earth's surface and atmosphere because more energy is being retained in the Earth system. SSP1-2.6 is often considered a stringent mitigation scenario where strong efforts are made to reduce greenhouse gas emissions while SSP5-8.5 is characterized by high greenhouse gas emissions, particularly carbon dioxide, resulting from continued fossil fuel use and limited mitigation efforts, and SSP2-4.5 is considered the middle of the road (Gidden *et al.*, 2019).

We select three different scenarios (SSP1-2.6, SSP2-4.5, SSP5-8.5) to quantify the sensitivity of vulnerability to food insecurity, and compute all variables, except hot days and wet days for 3 different years (2050, 2070 and 2090) across these three different scenarios (SSP1-2.6, 2-4.5 and 5-8.5). We select these three scenarios as they contrast in terms of their levels of radiative forcing, and their coupled social scenarios. We use the past 33 years (as we did when we constructed variables for 2013), where the construction of a variable relies on data prior to the year of interest.

Table 2 Climatic variables

Variable	Description	**Resolution**	**Year**
Mean monthly temperature	Average monthly air temperature in degrees Celsius, at 2 m above the surface, for the year 2013. Monthly averages are averages of daily air temperature.	~27km	2013
Temperature in first month of planting season	Average daily air temperature in degrees Celsius, at 2 m above the surface, in the planting season (Oct-Dec) in 2012.	~27km	2013
Variance of average monthly temperature	This variable is a measure of the variance of monthly average temperature (air temperature in degrees Celsius at 2 meters above the surface) across 1980 to 2013.	~27km	1980-2013
Variance of temperature in first months of planting season	This variable is a measure of the variance of monthly average temperature (air temperature in degrees Celsius at 2 meters above the surface) in the planting season (Oct-Dec) for the period 1980-2013.	~27km	1980-2013
Average monthly rainfall	Average monthly rainfall (in mm), for the year 2012. Monthly rainfall is calculated as the sum of daily sums.	~27km	2013

Variable	Description	**Resolution**	**Year**
Rainfall in first month of planting season	This variable is a measure of rainfall (in mm) in the planting season (Oct-Dec) in 2012.	~27km	2013
Variance of monthly rainfall	This variable is a measure of the variance of monthly rainfall (in mm) across 1980 to 2012.	~27km	1980-2013
Variance of rainfall in first month of planting season	This variable is a measure of the variance of monthly rainfall (in mm) in the planting season (Oct-Dec) for the period 1980-2012.	~27km	1980-2013
Standardised deviation of Temperature	This variable measures the deviation of total monthly rainfall in November or December 2012 from the average monthly rainfall between 1980 and 2013. The variable is standardised using the standard deviation of monthly rainfall between 1980 and 2012,	~27km	1980-2013
Standardised deviation of Rainfall	This variable measures the deviation of average monthly temperature in November or December 2012 from the average monthly rainfall between 1980 and 2013. The variable is standardised using the standard deviation of average monthly temperature between 1980 and 2012.	~27km	1980-2013

3. Methods

Our dataset contains the food consumption score, a set of natural capital and climate variables, and projections of climatic variables under different scenarios, for each household. This rich cross-sectional dataset enables us to analyse climatic drivers of vulnerability to food insecurity. We proceed in two steps: firstly, we estimate vulnerability to food insecurity for clusters of households defined on the basis of best predictors of food security, selected using machine learning techniques. Then, using these results, we predict changes to vulnerability to food insecurity to reflect different climate change scenarios.

3.1 Vulnerability to food insecurity

Estimating vulnerability to food insecurity required consistent estimates of the mean and variance of the FCS. In a regression context, the first step was to characterise household food consumption, c_h, as measured by the FCS, as a function of its observable characteristics, X_h, as

$$c_h = X_h\hat{\beta} + \mu_h \tag{2}$$

Using the estimates $\hat{\beta}$ we can estimate the expected food consumption score which, conditional on X_h, forms the deterministic component of the distribution of consumption:

$$\hat{E}[c_h|X_h] = X_h\hat{\beta} \tag{3}$$

and the variance of consumption, conditional on X_h:

$$\hat{V}[c_h|X_h] = \hat{\sigma}^2_{\mu,h} = X_h\hat{\theta} \tag{4}$$

for each household h. Assuming that the FCS is normally distributed, we can use these estimates to: (1) characterise the distribution of the FCS conditional of the set of covariates X_h and (2) estimate the probability that a household with characteristics X_h, would be food insecure – that is, estimate the household's vulnerability level. Letting Φ(.) denote the cumulative density function of the standard normal distribution, this estimated probability is given by:

$$\hat{v}_h = \widehat{Pr}(c_h < z|X_h) = \Phi\left(\frac{z - X_h\hat{\beta}}{\sqrt{X_h\hat{\theta}}}\right) \tag{5}$$

Empirically, a large literature in production economics, building on Just and Pope (1978, 1979) shows how to analyse the conditional variance of an outcome variable (in our case, food consumption score) as a function of observable characteristics of the household. Pritchett et al. (2000) and Chaudhuri et al. (2002) are two early examples of using a conceptually similar approach to estimate vulnerability to poverty. See Santos et al., (2023) for a review of this literature.

In this approach, we must first estimate the conditional mean of food consumption score as:

$$FCS_{ij} = \beta_0 + \beta_1 X_i + \beta_2 C_i + \varepsilon_{ij} \quad (6)$$

where FCS_{ij} is the food consumption score of household i living in community j, X_i are natural capital variables and C_i are climatic variables.

The estimates of the effect of observable characteristics on consumption is captured in β and ε_i is an idiosyncratic error term that captures the unobserved determinants of consumption. With these estimates, we can then obtain the empirical estimates of the first two moments of the distribution of FCS (ie, the equivalent to equations (3) and (4)) and obtain estimates of the effect of household covariates on the variance of consumption, in our case using the same set of covariates as in equation ():

$$\hat{\varepsilon}_i{}^2 = (FCS_i - E(FCS_i|X_i, S_i))^2 = \theta_0 + \theta_1 X_i + \theta_2 C_i + v_{ij} \quad (7)$$

Together with the estimates of the effect of each covariate on the expected value of FCS, we can then estimate the probability that each household will have a consumption below a specific threshold e estimated θ parameters allow us to quantify the main correlates of food consumption score risk and obtain estimates of vulnerability to food insecurity (in conjunction with the estimates of their effect on mean food consumption score).

3.2 Creating homogenous cohorts and predicting their spatial distribution

Implicit in the above framework is the assumption that, conditional on observable characteristics *X*, all observations fit the same food consumption score function. One way to avoid such a strong assumption is to use machine learning techniques to identify more homogeneous subgroups of observations.

The approach that we follow borrows from a related analysis, described in more detail in Santos et al (2023). Succinctly, we will focus on the use of regressions trees and related advances such as regression forests and the estimation of surrogate models, to re-estimate equation (6) as

$$FCS_i = \text{f1}\ (X_i, S_i,)\ if\ W \geq w0 \quad (8.1)$$

$$FCS_i = \text{f2}\ (X_i, S_i)\ otherwise \quad (8.2)$$

where the outcome depends on whether a specific variable *W* is above or below a certain cut-off (w_0). If the condition identified in equation (8.1) is met (or not), then the effect of other variables *(X, S)* is better expressed by function f_1 (or f_2, if not), rather than imposing a common functional relation as in equation (6). The statistical selection of variables *W* and their threshold levels, w_0, leads to the identification of a hierarchy of importance of those variables in predicting food consumption score.

3.2.1 Machine learning

A large (and growing) number of statistical approaches, under the label of machine learning, aim to capture the basic intuition underlying equations (8.1) and (8.2): that it is better (in a predictive sense) to account for heterogeneity rather than assume homogeneity. This improvement in predictive power comes at the cost of increased complexity. The model captured by equations (8.1) and (8.2) is less parsimonious than the one described by equation (6), and this added complexity needs to be (negatively) weighted against the gain in predictive accuracy.

The main reason for using machine learning algorithms is to create a relatively strong predictive model (even if it is a black box model), to predict the food consumption score under different climate scenarios as accurately as possible. As in Santos et al (2023) we use a surrogate decision tree model based on the predictions of a random forest model to produce a simpler and interpretable model of the role played by different covariates in determining our predictions of the impact of climate change, and in the process, create cohorts as units for further analysis.

4. Results

We present 3 main sets of results.

1. analysis of the importance of environmental variables as predictors of food consumption score, which are made interpretable using surrogate models.
2. vulnerability to food security, including different results by the clusters formed by the surrogate model
3. predictions of food insecurity for households and clusters under different climate scenarios.

4.1 Prediction of food insecurity

We estimate a random forest model on a training set (60% of the sample, used to tune the different parameters of each model) and assess its accuracy using key performance metrics such as Rsquared and the RMSE as evaluated in the test set (40% of the sample). We run several different specifications to understand the performance of the model (see Table 3). We find little difference in predictive power across our models, and hence use a model that addresses the motivation of this study (to quantify the relation between food security and climatic variables under different scenarios) more directly (in bold, in table 3).

Table 3- Rsquared and RMSE, random forest specification

	RSquared	RMSE
Random Forest – all environmental variables	0.141	14.9
Random forest – climatic variables	**0.138**	**14.9**
Random forest - clusters	0.131	14.9

Random forest algorithms are complex and are not interpretable (when compared, for example, with OLS regression or regression trees), making it impossible to identify the main predictors of food insecurity. We overcome this limitation by using surrogate models to understand which environmental variables are most important in predicting consumption.

Figure 3 presents the results of this approach. The R-squared for the surrogate model relative to the random forest model is .316. In this model, differences in average temperature in the first two months of the main planting season appear to have the greatest influence in terms of organizing the data: the full sample is first split into 2 groups as a function of whether the realized value of this variable is above a threshold of 25C, with warmer areas in those two months registering better outcomes in terms of average food security (and appearing to the left of this first split in the surrogate tree). Further splits of the sub-samples, as a function of whether the different criteria are met (and the observations show up to the left of the node) or not (observations are clustered to the right of the node) then allows us to present a clearer picture of the influence of climatic variables on the food consumption score (and food security). This logic can be followed until reaching the leaf (final) nodes.

We can make some observations from the leaf nodes:

- First, the regression tree succeeds in creating groups with meaningful differences in consumption: the expected food security in the most deprived cluster is 31 which is 14 points below the cluster with highest expected food consumption score.
- Second, temperature seems to matter most among the weather conditions, either in terms of the average monthly temperature at the start of the planting season, and the variance of the average monthly temperature in the planting season. Higher average monthly temperatures in the planting season and less variance in average monthly temperatures in the planting season (lower risk) are linked to higher expected food consumption scores. However, and importantly, this relation is nonlinear: households living in areas with unusually high average temperatures at the start of the planting season (>29C) are much better off (in terms of average diet quality) than 90% of the rural households in the country.
- Third, natural capital (terrain characteristics and soil properties) do not seem to matter in terms of predicting homogeneous clusters: instead, all predictors are climatic variables, highlighting the importance of global warming for these outcomes in Mozambique.
- Finally, highlighting the predictive power of this model, the (weighted) frequency of food insecurity (55%) if essentially identical to the unconditional probability of being food insecure (shown in figure 1).

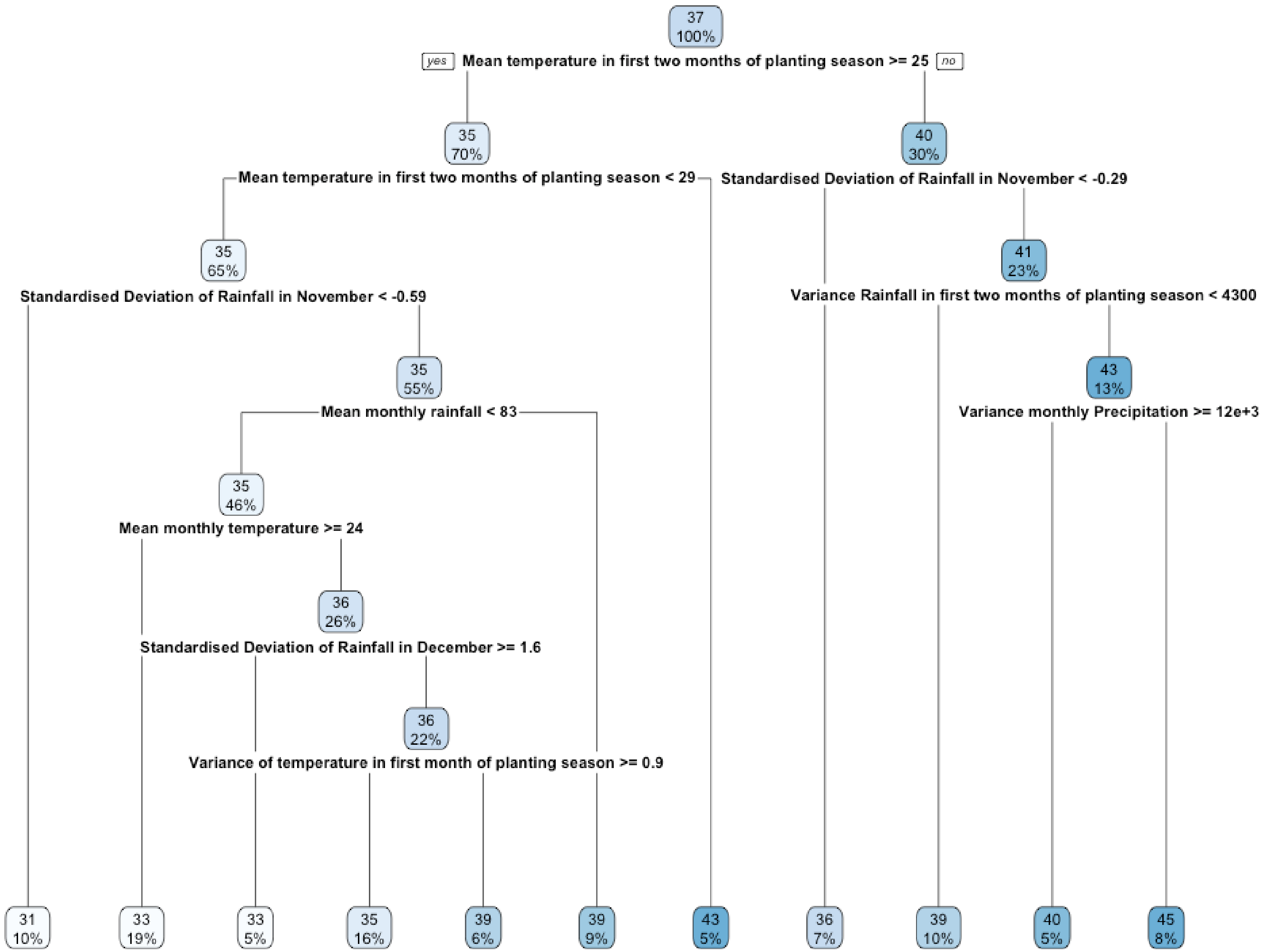


Figure 3 Surrogate model

4.2 Characterising food insecurity

The surrogate model created 11 groups (corresponding to the final leaf nodes), characterized in terms of differences in expected food consumption score (our outcome of interest) and its set of predictors, selected among a larger set that includes both natural capital and climatic variables. Figure *4* shows the distribution of these groups across the territory, with purple colours showing lower predicted food consumption scores (higher food insecurity) and green to yellow colours showing higher food consumption scores. There are some geographic groupings of the different cohorts. For the most part, households in the higher groups (higher FCS) are located in the centre-west of the country, and households in lower groups (lower FCS) are located in the centre-east of the country. We also present the data in Figure 5, Figure 6 and Figure 7 by the splits in temperature that are observed at the top of Figure *4*.

Table 3 presents the average characteristics of the households in each cohort. These results suggest the following comments. Firstly, and as expected, our model does a fairly good job at tracking average diet quality: average predicted FCS closely resembles observed values of this indicator. However, its performance is lower when focusing on predicting food insecurity (ie, FCS<=35), particularly in those cohorts that are, on average predicted to be better-off (ie, cohorts 6 to 11). Future work may re-examine this feature of the analysis by explicitly modelling food insecurity (an indicator variable) as the outcome of this model instead of FCS (a continuous variable).

Secondly, households in different cohorts do not seem substantively different in terms of demographic characteristics, including dependency ratio and, to a lesser degree, number of adults (as proxy for access to labor). However, they do seem to make different use of migration as a livelihood strategy, which differs significantly between cohorts and seems mostly associated with lower average FCS, suggesting its use as a coping strategy.

Thirdly, and reflecting the sampling criteria of the IAI2012, households do not differ in terms of access to land (all households have less than 3 hectares of land, regardless of their cohort). Future work may want to explore the performance of larger farms, making using of their oversampling in the most recent wave of the IAI. They do differ, however, in their use of land, most significantly in the importance of the area devoted to pasture. These differences are not mimicked in the importance of ownership of cattle: the data suggests that some the better-off cohort associated with higher temperatures at the start of the production season (>29C) may be better characterized by intensive cattle production, given their large ownership of cattle (>5 animals) and almost no land devoted to pasture or fallow.

Finally, the exposure of different cohorts to natural hazards and shocks is also different: while the prevalence of fires (and also pests and damages by wildlife, not shown in table 3 for simplicity) is fairly similar across the entire sample, the same is not true for floods and droughts. Although these shocks may partially reflect their membership in specific cohorts (themselves defined on the basis of climatic variables), future work may explore the importance of such extreme events in explaining (in)security.

A more complete characterization of the household's vulnerability to food insecurity requires more than an analysis of average outcomes including, as made clear by equation (4), a characterization of conditional variance. We can then combine the first two moments of the distribution of FCS to quantify vulnerability (equation (5)). Figure 8 and Figure 9 present the results of this analysis for all households and for each of the 11 groups, respectively, where the y-axis is the percentage of observations in the sample/group and the x-axis represents the probability that the household will fall below the food insecurity line. As Figure 9 makes clear, there are different distributions of vulnerability to food insecurity for each group that reflect more than differences in mean FCS.

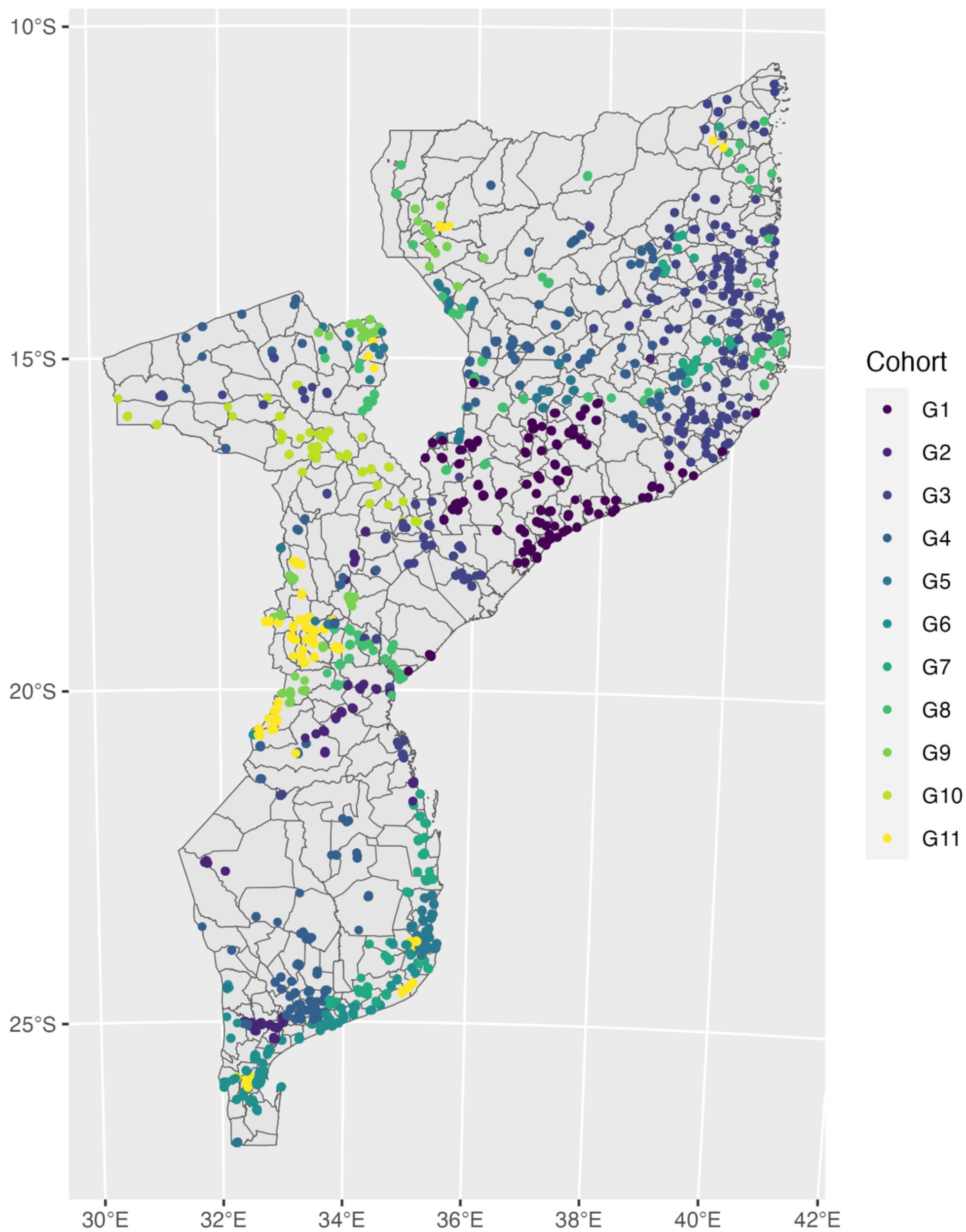


Figure 4 Household cohort, by location (Note: exact locations masked for this figure)

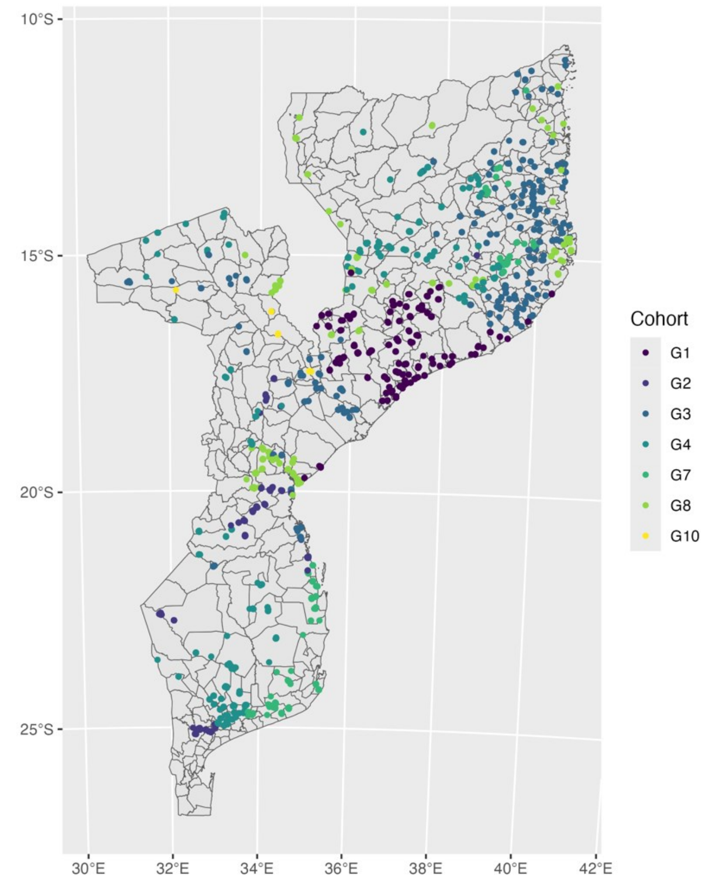


Figure 5 Household cohort, by location (Note: exact locations masked for this figure): filtered for households where the average temperature for November is <= 29 and > 25 degrees Celsius

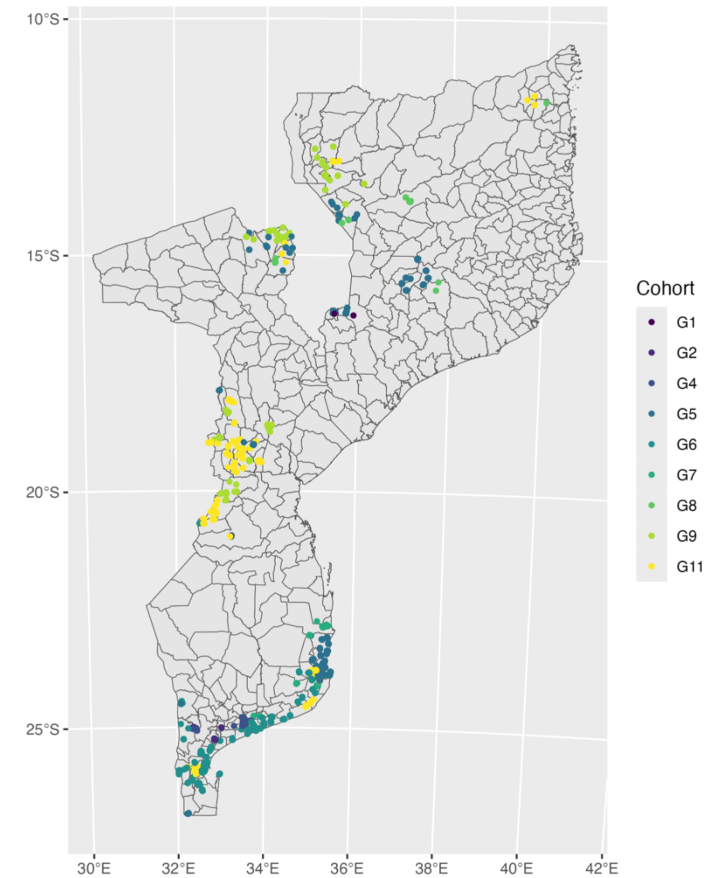


Figure 6 Household cohort, by location (Note: exact locations masked for this figure): filtered for households where the average temperature for November is < 25 degrees Celsius

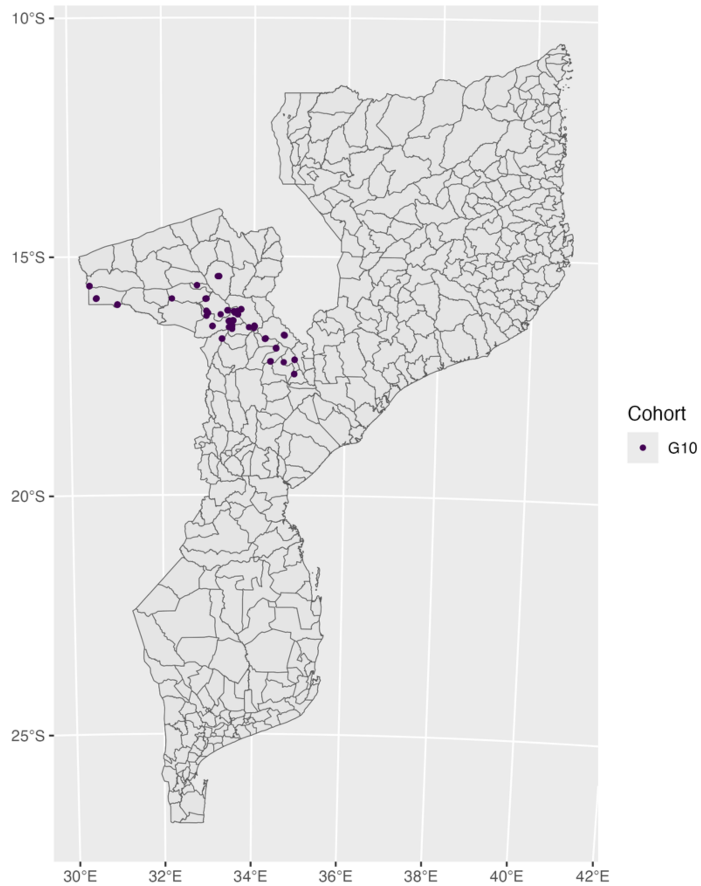


Figure 7 Household cohort, by location (Note: exact locations masked for this figure): filtered for households where the average temperature for November is > 29 degrees Celsius

Table 3 Average characteristics of households in different cohorts

Variable Cluster ->	1	2	3	4	5	6	7	8	9	10	11
Mean Food consumption score	30.03	32.50	33.04	35.46	36.58	39.33	38.58	38.88	40.34	41.32	45.54
predicted food consumption score	30.77	32.82	33.40	35.30	36.32	38.83	39.00	38.98	39.89	42.73	44.80
proportion of households food insecure	0.71	0.66	0.63	0.57	0.55	0.49	0.48	0.47	0.41	0.43	0.34
predicted proportion of households food insecure	0.85	0.72	0.69	0.51	0.45	0.26	0.29	0.30	0.24	0.14	0.07
Proportion of households where household head has above high school education	0.00	0.01	0.00	0.00	0.00	0.03	0.01	0.01	0.01	0.01	0.02
Proportion of households where household head is a farmer	0.81	0.69	0.89	0.81	0.76	0.54	0.74	0.78	0.80	0.82	0.63
Proportion of households where household head is absent for 3+ months	0.07	0.06	0.02	0.08	0.04	0.09	0.07	0.03	0.02	0.03	0.04
Number of adults	2.08	2.96	2.20	2.72	2.32	2.82	2.40	2.51	2.42	2.51	2.85
Dependency ratio	1.42	1.30	1.36	1.46	1.27	1.19	1.35	1.32	1.40	1.43	1.26
Migrants	0.18	0.24	0.05	0.31	0.13	0.24	0.25	0.10	0.05	0.09	0.18
land (ha)	1.79	2.88	2.30	2.76	1.83	2.03	1.91	2.02	2.25	2.01	2.41
crops (ha)	1.63	2.51	1.88	2.49	1.68	1.51	1.67	1.82	2.09	1.97	1.91
fallow (ha)	0.13	0.30	0.13	0.23	0.11	0.44	0.17	0.15	0.14	0.05	0.22
pasture (ha)	0.02	0.96	0.07	0.25	0.27	1.77	0.15	0.01	0.02	0.01	0.07
cattle	0.03	4.08	0.60	3.93	0.99	3.16	0.58	0.47	0.65	5.44	1.04
fire	0.15	0.10	0.08	0.10	0.10	0.11	0.09	0.10	0.11	0.04	0.10
drought	0.45	0.37	0.18	0.47	0.45	0.48	0.58	0.21	0.23	0.49	0.42
floods	0.33	0.33	0.16	0.22	0.10	0.23	0.06	0.21	0.10	0.18	0.10

Figure 8 Probability of food insecurity

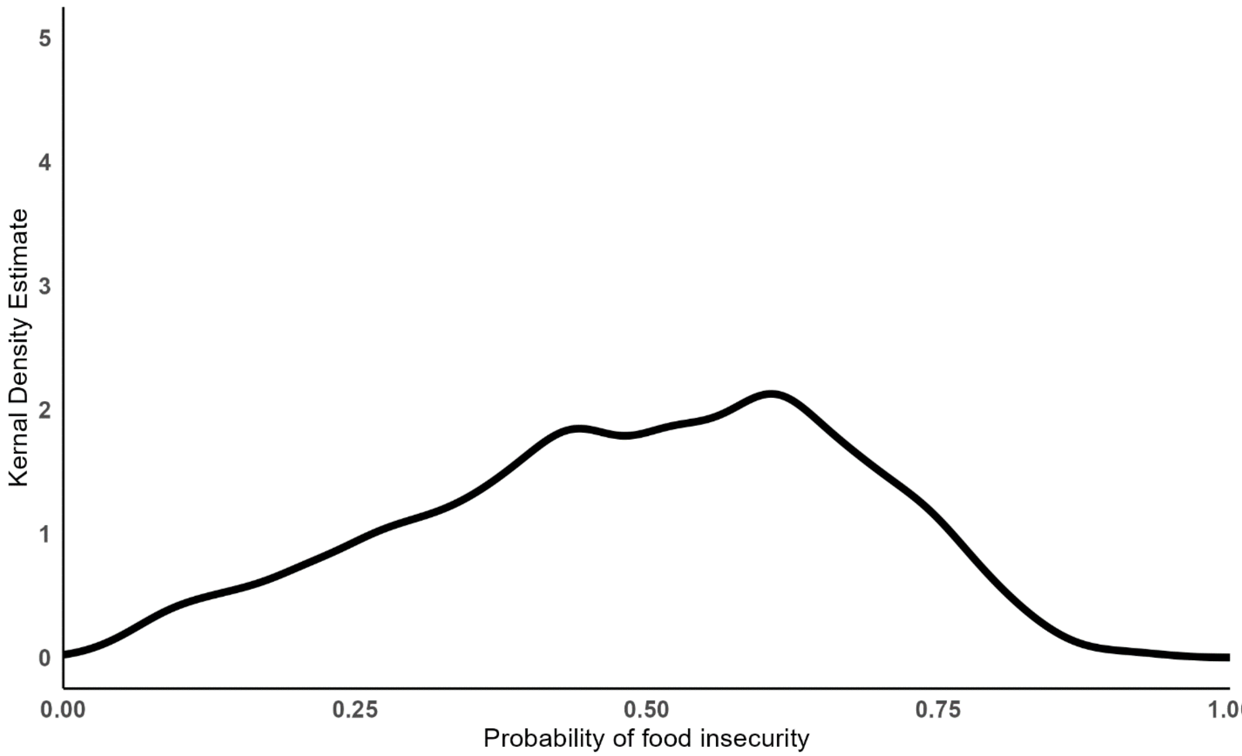
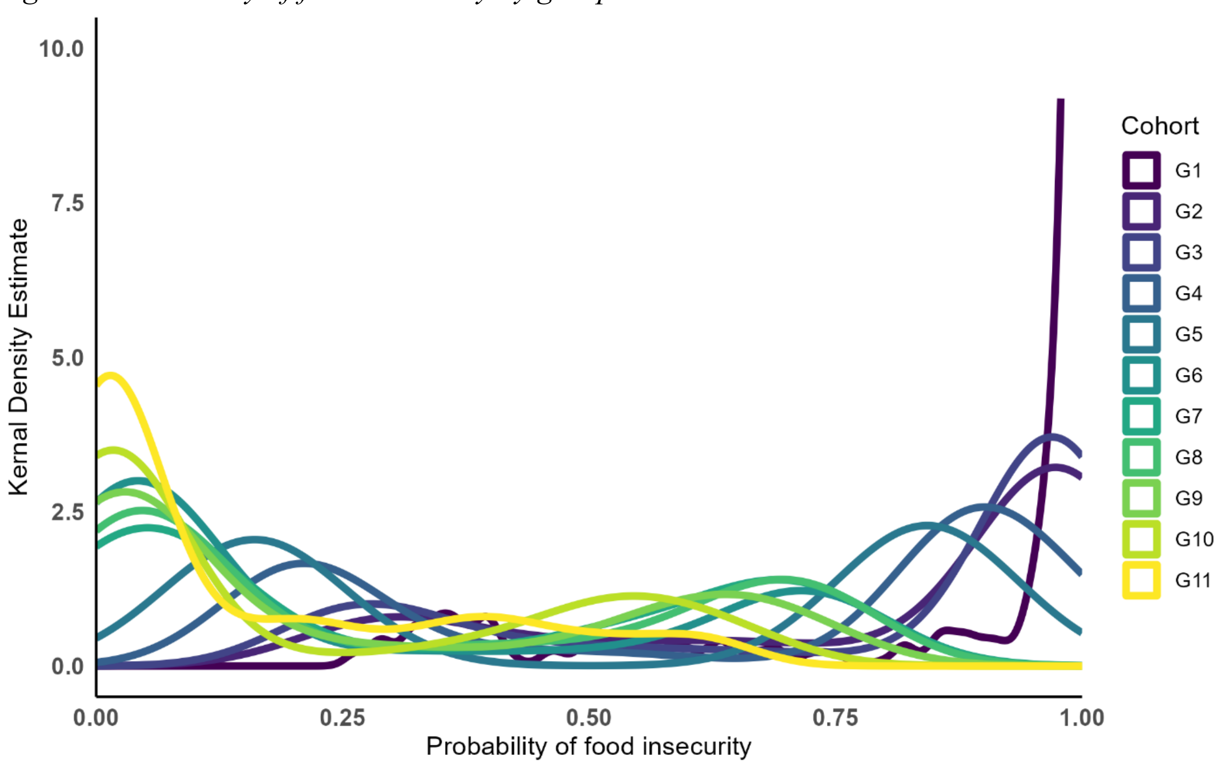


Figure 9 Probability of food insecurity by group



NB: using <36 as threshold

4.3 Predicting food insecurity under different climate change scenarios

Predicting future food insecurity based on climate scenarios is a relatively straightforward exercise: it requires substituting our climatic variables in equation 5 with the respective values under different scenarios (see the discussion in section 2.1.2.2 on SSP 1-2.6, SSP 2-4.5 and SSP 5-8.5) and using the decision tree (figure 4) to determine their (potentially, new) group. We use our random forest model to predict the changes in the FCS under different climate scenarios, and under two possibility set of livelihood strategies: first, when the set of livelihood strategies adopted by households in the warmest locations (cluster 10, in Table 3) is not present (and, consequently, the negative relation between temperature at the start of the main agricultural season and diet quality is monotonic), which we label as "adaptation: restricted" in the table below; and when those activities are feasible (and the relation between temperature and diet quality is nonlinear, as presented in Figure 4), which we label below as "adaptation: unrestricted".

Table 4: Simulating the effect of global warming on food security, under different adaptation possibilities

	Adaptation: restricted			Adaptation: unrestricted		
scenario	2050	2070	2090	2050	2070	2090
SSP 1-2.6	34.52 (4.47)	38.24 (2.58)	36.57 (3.73)	35.64 (4.91)	41.73 (1.75)	41.36 (2.47)
SSP 2-4.5	37.97 (2.70)	37.16 (3.63)	38.96 (1.74)	41.29 (2.60)	41.87 (1.66)	40.83 (1.87)
SSP 5-8.5	34.52 (4.18)	34.93 (3.86)	36.29 (3.81)	41.81 (1.71)	42.67 (0.98)	42.67 (0.98)

The conclusions from this analysis are relatively straightforward. Firstly, if the set of livelihood activities currently practiced by households in warmer locations (>29C, cluster 10 in Table 3) is not available, the dismal levels of household food security in rural Mozambique observed in the IAI2012 are expected to remain unchanged. However, their variability is expected to be smaller. Secondly, if adaptation is unrestricted, changes associated with global warming would lead to improvements in diet quality and associated reduction in food insecurity.

This result is perhaps unexpected and the analysis of the surrogate model enables us to understand what is driving these changes in FCS (and the changes in vulnerability profiles): for the most part, these changes reflect the fact that, on the basis of the data of IAI2012, higher temperatures (ie,

>29C in the first two months of the main production season) predict higher average FCS in cluster 10 (in table 3). Implicit in this “movement” is the assumption that adaptation costs in terms of switching between groups are not prohibitive. As argued above, the analysis of table 3 suggests that this assumption may be fairly plausible in some cases but likely too strong in others (eg, when it requires a switch to what is an apparent focus on intensive cattle production, as practiced by households in warmer areas).

5. Conclusion

This report provides a first attempt at quantifying the effect of global warming on vulnerability to food insecurity in rural Mozambique. It combines detailed household data with a wide array of environmental datasets to characterize the plausibly heterogeneous relation between household wellbeing and the environment, which is quantified using machine learning techniques (regression forests and associated surrogate models). Two main conclusions emerge: firstly, that vulnerability to food insecurity in rural Mozambique among smallholders is high (as expected) but heterogeneous. This heterogeneity seems to reflect position in the territory and, to a smaller extent, difference in production decisions. Secondly, the impacts of global warming on vulnerability to food insecurity will, ultimately, depend on whether adaptation is feasible given the set of constraints faced by rural households. The analysis of the relation between food security and local environment based on the IAI2012 suggests that there is some scope for adaptation to higher temperatures.

These results, and in particular the perhaps unexpected prediction of reductions in vulnerability to food insecurity, require further analysis. Most immediately, and as already noticed, this conclusion assumes that adaptation costs to changes in rainfall and temperature patterns are minimal. This is obviously incorrect, but suggests an important role for policy in minimizing those costs, and where in the territory they may matter most.

Proceeding with the policy implications, it is important to notice that we analyzed the impact of climate change on diet quality, as a summary measure of household wellbeing. However, this focus ignores the multiplicity of livelihood strategies (crops grown, animals raised, investment in other livelihood strategies such as temporary migration) that underly this outcome. A necessary next step in the definition of policies that facilitate adaptation to climate change is to dig deeper into those choices. The IAI provides enough detail for that future analysis.

Finally, it is important to notice that the strength of this result may reflects two important limitations of our analysis. The first is that we rely only on one wave of the IAI. Although large cross-sections (such as IAI2012) are typically seen as adequate to estimate vulnerability in past work that we follow, that same literature also makes the obvious point that variability in time can only strengthen the conclusions of this type of analysis. Extending the analysis to the IAI2020 is an obvious next step in this analysis. The second limitation, that again may reflect the fact that we only use one cross-section of household data, is that shocks (drought, floods, etc) do not show up as important determinants of heterogeneity in household wellbeing. To the extent that this result may reflect the environmental conditions in the period of household data collection (and the fact that the frequency of these shocks seems to fairly homogeneous across groups), it may also underestimate the impact of climate change (and associated increased frequency of these shocks) on our results.

References

Chaudhuri, S., Jalan, J. and Suryahadi, A. (2002) ‘Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia’, *World*, 0102–52(April), p. 36.

Copernicus Climate Change Service (2021) ‘CMIP6 predictions underpinning the C3S decadal prediction prototypes’. ECMWF. Available at: https://doi.org/10.24381/CDS.C866074C.

Copernicus Climate Change Service (2023) ‘Complete ERA5 global atmospheric reanalysis’. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available at: https://doi.org/10.24381/CDS.143582CF.

Gidden, M.J. *et al.* (2019) ‘Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century’, *Geoscientific Model Development*, 12(4), pp. 1443–1475. Available at: https://doi.org/10.5194/gmd-12-1443-2019.

Hengl, T. (2018) ‘Soil pH in H2O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution’. Zenodo. Available at: https://doi.org/10.5281/ZENODO.2525664.

Hengl, T. and Ichsani Wheeler (2018) ‘Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution’. Zenodo. Available at: https://doi.org/10.5281/ZENODO.2525553.

Just, R.E. and Pope, R.D. (1978) ‘Stochastic specification of production functions and economic implications’, *Journal of Econometrics*, 7(1), pp. 67–86. Available at: https://doi.org/10.1016/0304-4076(78)90006-4.

Just, R.E. and Pope, R.D. (1979) ‘Production Function Estimation and Related Risk Considerations’, *American Journal of Agricultural Economics*, 61(2), pp. 276–284. Available at: https://doi.org/10.2307/1239732.

Nunn, N. and Puga, D. (2012) ‘Ruggedness: The Blessing of Bad Geography in Africa’, *The Review of Economics and Statistics*, 94(1), pp. 20–36.

Pritchett, L., Suryahadi, A. and Sumarto, S. (2000) ‘Quantifying vulnerability to poverty - a proposed measure, applied to Indonesia’, *Policy Research Working Paper Series* [Preprint]. Available at: https://ideas.repec.org//p/wbk/wbrwps/2437.html (Accessed: 6 June 2024).

Riley, S.J., De Gloria, S.D. and Elliot, R. (1999) ‘A Terrain Ruggedness Index that Quantified Topographic Heterogeneity’, *Intermountain Journal of Sciences*, 5(1–4), p. 23:27.

Santos, P. *et al.* (2023) *Vulnerability in the Anthropocene: a prospective analysis of the need for social protection*. Australian Centre for International Agricultural Research. Available at: https://www.aciar.gov.au/publication/TR102-vulnerability (Accessed: 6 June 2024).

Yamazaki, D. *et al.* (2017) 'A high-accuracy map of global terrain elevations', *Geophysical Research Letters*, 44(11), pp. 5844–5853. Available at: https://doi.org/10.1002/2017GL072874.

IGC

theigc.org