Methods Paper Using Household survey data to identify large-scale food security patterns across Uganda

PLOS One

Jannike Wichern[1], Joost van Heerwaarden[1], Mark van Wijk[2]
[1]Wageningen University, [2]International Livestock Research Institute

dx.doi.org/10.17504/protocols.io.vb6e2re

Jannike Wichern

ABSTRACT

To target food security interventions for smallholder households, decision makers need large-scale information, such as maps on poverty, food security and key livelihood activities. Such information is often based on expert knowledge or aggregated data, despite the fact that food security and poverty are driven largely by processes at the household level. At present, it is unclear if and how household level information can contribute to the spatial prediction of such welfare indicators or to what extent local variability is ignored by current mapping efforts. A combination of geo-referenced household level information with spatially continuous information is an underused approach to quantify local and large-scale variation, while it can provide a direct estimate of the variability of welfare indicators at the most relevant scale. We applied a stepwise regression kriging procedure to translate point information to spatially explicit patterns and create country-wide predictions with associated uncertainty estimates for indicators on food availability and related livelihood activities using household survey data from Uganda. With few exceptions, predictions of the indicators were weak, highlighting the difficulty in capturing variability at larger scale. Household explanatory variables identified little additional variation compared to environmental explanatory variables alone. Spatial predictability was strongest for indicators whose distribution was determined by environmental gradients. In contrast, indicators of crops that were more ubiquitously present across agroecological zones showed large local variation, which often overruled large-scale patterns.

Our procedure adds to existing approaches that often only show large-scale patterns by revealing that local variation in welfare is large. Interventions that aim to target the poor must recognise that diversity in livelihood activities for income generation *within* any given area often overrides the variability of livelihood activities between distant regions in the country.

EXTERNAL LINK
https://doi.org/10.1371/journal.pone.0208714

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

to be added

PROTOCOL STATUS

Working

GUIDELINES

23 October 2018

Information:
This is the analysis that belongs to the publication Wichern et al.: Using household survey data to identify large-scale food security patterns across Uganda, to be published in PlosOne

Data can be downloaded from the following sources:

1. For the Food availability analysis:
LSMS household survey data 2010/11 for Uganda: http://microdata.worldbank.org/index.php/catalog/2166
The scripts have been adjusted based on Frelat et al. 2016 (PNAS): https://doi.org/10.1073/pnas.1518384112

2. For the spatial analysis:
Climate data: Hijmans et al. 2005, BioClim data version 1.4, http://www.worldclim.org/
Elevation: Jarvis et al. 2008, http://www.cgiar-csi.org/data/

Soil data: Hengl et al. 2017
Population:Worldpop 2015, Worldpop.org.uk
Market access: Nelson 2008, http://forobs.jrc.ec.europa.eu/products/gam/download.php
Length of growing period: HarvestChoice 2015, Harvestchoice.org

BEFORE STARTING

The analysis is performed in R. Make sure you have all the required packages installed (see scripts).

Spatial interpolation of household survey data

1 Step 1: Household level food availability analysis

In this step the household data are used to calculate food availability and the contributing activities on household level.
To run the household level food availability analysis household survey data must be downloaded from the World Bank for 2010/11 for Uganda: http://microdata.worldbank.org/index.php/catalog/2166

Run the household level food availability analysis (R-scripts attached, scripts 1.1. to 1.5). Make sure that the file sources are adjusted in the scripts LSMS_Tools.R and LSMS_201011_param.R to be able to run script 1.4_LSMS_201011_IA.R. Two csv-files attached provide parameters in LSMS_201011_param.R.

1.1_LSMS201011_HHinfo_281016.R
1.2_LSMS_201011_Crop_281016.R
1.2a_LSMS_201011_Calculate_median_cropPrices_281016.R
1.3_LSMS_201011_Lvst_281016.R
1.3a_LSMS_201011_Calculate_median_livestockPrices_281016.R
1.4_LSMS_201011_IA_281016.R
1.5_LSMS_201011_Calc_contribution_crop_lvst_281016.R
LSMS_201011_param_281016.R
LSMS_Tools_281016.R
LSMS_Uganda_CropEnergy_3.0_LSMS201011.csv
LSMS_Uganda_LvstEnergy_3.0_LSMS201011_FAO.csv

2 Step 2: Data preparation for regression analysis

In this step the household data and spatial data layers are prepared for the regression analysis.
First, the LSMS household locations are randomly off-set (step 2.1.). Next, the soil carbon stock is calculated (2.2.). Then the spatial layers are sampled to the off-set household locations (step 2.3.).
For step 2 the household output data from step 1 are needed as well as the following spatial data:

Climate data: Hijmans et al. 2005, BioClim data version 1.4, http://www.worldclim.org/
Elevation: Jarvis et al. 2008, http://www.cgiar-csi.org/data/
Soil data: Hengl et al. 2017
Population:Worldpop 2015, Worldpop.org.uk
Market access: Nelson 2008, http://forobs.jrc.ec.europa.eu/products/gam/download.php
Length of growing period: HarvestChoice 2015, Harvestchoice.org

Details on the sources are in the paper Wichern et al. in Plos One.

2.1.Random_coordinate_offset_081116.R
2.2.soil_carbon_stock.R
2.3.Resampling_res_sample_spatial_to_point_5arcmin_180611.R

3 Step 3: Regression analysis

For each variable of interest (see paper Wichern et al. in Plos One) regression models are built using the spatial variables from step 2 and household variables from step 1. Depending on the characteristics of the variable of interest, a multiple linear regression model (MLR, step 3.2.) or a multiple inflated beta regression model (MIBR, step 3.1.) is used. Scripts here are examples for the variables banana contribution and food availability.

3.1.MIBR_Banana_5arcmin_180612.R
3.2.MLR_FoodAvail_5arcmin_180611.R

4 Step 4: Spatial interpolation

Once the regression model is built, the model is used to spatially interpolate the variable of interest. Depending on the regression model type, the spatial interpolation either is done one a multiple linear regression model (4.2.) or on a multiple inflated beta regression model (step 4.1.).

4.1.MIBR_Banana_PART2_5arcmin_180612.R
4.2.MLR_FoodAvail_PART2_5arcmin_180611.R