Documentation SIMLESA 2013

Romain Frelat

27-05-2024

Objectives

This document records the steps and choices made to transform SIMLESA Malawi 2013 dataset into the format of the Household Data Platform. It goes together and complements the script file `Get_SIMLESA_MWI2013.R`. The original dataset is:

Pathways to sustainable intensification in Eastern and Southern Africa - Malawi 2013 Marenya, Paswel; Kassie, Menale; Muricho, Geoffrey; Alemu, Solomon; Yirga, Chilot; Mishili, Fulgence; Obare, Gideon; Tostao, Emilio, 2016, "Pathways to sustainable intensification in Eastern and Southern Africa - Malawi 2013", https://hdl.handle.net/11529/10760, CIMMYT Research Data & Software Repository Network, V3, UNF:5:ascjPZI1bFZMaMFDIMe7uA

Crop table

The crop table contains 7 columns with information on crop cultivation and usage.

Name	Definition	Unit
hhid	household id	
name	name of the crop	
land_area_ha	land cultivated	hectare
harvest_kg	amount harvested	kg
consumed_kg	amount consumed	kg
sold_kg	amount sold	kg
income_lcu	income from sells	lcu

The land and crop management are described in Module 4 Part A, the utilization of crop harvested in Module 4 Part B, and the incomes in Module 4 Part C.

Land area

We estimated the land area per crop from the questions in module 4, part A.

The field size were already provided in hectare. We used the reported percentage of the sub-plot cover by the three crops to estimate the area per crop. We corrected these percentages when necessary to make sure that the sum of the crop areas per sub-plot is not higher than the sub-plot area. When the percentage was not reported, we estimated it based on the ratio of yields among the three main crops of the sub-plot. The remaining area (if any) was attributed to tree crop planted on the same field. Yet in most cases the tree crop (mostly banana and coffee) don´t have proper land area (nor information about harvested quantities).

Harvest and uses

The quantities of crop harvested are provided both in module 4 part A and part B. We decided to use the quantity of crop harvested provided in part B in priority (unless it was null, in this case we used the quantity

reported in part A), because it was easily related to crop uses. Similarly for the quantity of crop sold is provided in module 4 part B and C and we decided to use the quantity provided in part B in priority.

At the end, the quantities were checked for consistency. When the amount consumed and sold was higher than the amount harvested, we used the sum of the amount consumed and sold as quantity harvested, unless the resulting yield was higher than 20t/ha. In such case (abnormally high quantities sold and consumed), we lower the quantities sold and consumed using the same proportions as originally reported but calculated on the amount harvested.

Livestock table

The livestock table contains 3 columns with information on livestock herd

Name	Definition	Unit
hhid	household id	
name	name of the livestock	
n	number of livestock kept	

The information on the herd structure is provided in Module 10 part E: Livestock ownership.
We grouped *Crossbred/exotic*, *Heifers*, *Bulls*, *Calves*, *Oxen*, *Indigenous cows* into **cattle**.
We grouped *Donkeys*, *Horses*, and *Mule* into **Donkeys_horses**.

Livestock production table

The livestock production table contains 7 columns with information on livestock productions:

Name	Definition	Unit
hhid	household id	
name	name of the livestock	
prod	livestock production	
harvest_kg	amount harvested	kg
consumed_kg	amount consumed	kg
sold_kg	amount sold	kg
income_lcu	income from sells	lcu

The information on milk, dairy products, eggs and honey are captured in Module 10 part E: Livestock ownership.

The quantities consumed or harvest was not reported. So we used the quantities sold as quantity harvested and the quantities consumed was set to 0.

The income was calculated as the product of the quantity sold (A6) and the price per kg (A7). We checked these values by calculating the median price per item per kg, and we used the reported price, unless it was too different than the median price (=5 times higher or lower) and the value reported in `valsold` was closer to the median price.

The quantities sold (in numbers) are transformed in kg using the conversion of one egg to 0.05 kg, and the livestock sold alive using the TLU conversion factors (1TLU=250kg).

```
##        beehive         cattle   donkey_horse            pig smalllivestock
##            0.0          175.0          175.0           75.0           25.0
##        poultry
##            2.5
```

Household information table

The household information table contains information on household composition, off farm activities, and food security.

Name	Definition	Unit
hhid	household id	
country	country of the survey	
year	year of the survey	
gps_lat	latitude in decimal degrees	°N
gps_lon	longitude in decimal degrees	°E
hh_size_members	size of the household in number of persons	
hh_size_mae	size of the household in male adult equivalent	MAE
head_age	age of the household head	
head_gender	gender of the household head	'f' or 'm'
off_farm_lcu	off farm income per year	lcu
off_farm_div	diversity of off farm activities	
hdds	household diet diversity score based on 10 groups	
fies	Food Insecurity Experience Scale based on 8 questions	
foodshortage_count	number of months with food shortage	
foodshortage_months	name of the months with food shortage	
currency_conversion	conversion from local currency to power parity purchase usd	lcu/usd

General information

`hhid`is made of the prefix "SIMLESA_MWI_2013" and the variable `hhid`. This variable is a unique identifier for the household that can connect all the tables (crop, livestock, and livestock production).

The GPS coordinates were not provided. Therefore, we used the centroid of the reported district or EPA as coordinates of the households.

The currency power parity purchase conversion factor is provided by the World Bank for Malawi in 2013 (= 114.54mwk/usd).

Household size

The household composition is provided in Module 2 part A. The male adult equivalent are calculated with 5 age-groups: 0-4, 5-10, 11-24, 25-50, 51+ and with the corresponding coefficient per gender and age class. We used the reported age in month when the age in years was missing (for babies).

```
##        0-4 5-10 11-24 25-50  51+
## Male   0.5 0.75 0.925  1.00 0.73
## Female 0.5 0.75 0.750  0.86 0.60
```

Off farm income

We considered the off farm income reported in Module 11: Other sources of household income.

Food security

The diet diversity was calculated from the 7-day recall of food consumption (module 8). We grouped the 18 groups into 10 groups (to be similar to the LSMS-ISA Malawi survey).

```
##       Original                                Merged
##  [1,] "Cereals"                               "G1_Grains_Roots_Tubers"
##  [2,] "Vitamin A rich vegetables and tubers" "G8_VitA_Fruits_Vegetables"
##  [3,] "White tubers and roots"                "G1_Grains_Roots_Tubers"
##  [4,] "Dark green leafy vegetables"           "G7_Leafy_Vegetables"
##  [5,] "Other vegetables"                      "G9_Other_Vegetables"
##  [6,] "Vitamin A rich fruits"                 "G8_VitA_Fruits_Vegetables"
##  [7,] "Other fruits"                          "G10_Other_Fruits"
##  [8,] "Organ meat (iron rich)"                "G5_Meat"
##  [9,] "Flesh meats"                           "G5_Meat"
## [10,] "Eggs"                                  "G6_Eggs"
## [11,] "Fish"                                  "G5_Meat"
## [12,] "Legumes, nuts and seeds"               "G3_Nuts_Seeds"
## [13,] "Milk and milk products"                "G4_Milk"
## [14,] "Oils and fats"                         "G11_Fats Oil"
## [15,] "Red palm products"                     "G1_Grains_Roots_Tubers"
## [16,] "Sweets"                                "G12_Sugar"
## [17,] "Spices, condimets, beverages"          "G13_Spices"
## [18,] "Meal outside home"                     NA
```

Crop and Livestock summary

The household information table also contains 28 columns with summary information from crop and livestock tables.

Name	Definition	Unit
hhid	household id	
land_cultivated_ha	total land cultivated	ha
crop_div	number of crop cultivated	
crop_name	names of crop cultivated	
crop_harvest_kg	total crop harvest	kg
crop_yield_kg_per_ha	crop yield	kg/ha
crop_sold_kg	quantity of crop sold	kg
crop_sold_perc	percentage of quantities of crop sold	%
crop_income_div	number of different crop sold	
crop_income_lcu	total income from crop production	lcu
crop_value_lcu	value of crop produced but not sold	lcu
crop_consumed_kcal	energy value from crop consumed	kcal
livestock_tlu	herd size	tlu
lstk_div	number of livestock species herded	
lstk_name	names of livestock species herded	
lstk_harvest_kg	total livestock product harvested	kg
lstk_sold_kg	quantity of livestock product sold	kg
lstk_sold_perc	percentage of livestock production sold	%
lstk_income_div	number of different livestock products sold	
lstk_income_lcu	total income from livestock production	lcu
lstk_value_lcu	value of livestock production not sold	lcu
lstk_consumed_kcal	energy value from livestock consumed	kcal
farm_div	number of crop and livestock species	
farm_harvest_kg	total farm production	kg
farm_sold_perc_kg	percentage of farm production sold	%
farm_income_div	number of different farm products sold	
farm_income_lcu	total income from farm production	lcu
tot_income_lcu	total income (farm + off farm)	lcu
farm_consumed_kcal	energy value from farm production consumed	kcal

All the values are calculated automatically with the function `calc_farm_prod()`. The calculations are simple and summarize per households the quantities reported in the crop and livestock tables.

For energy conversion, we used estimates mostly from the FoodData Central of the U.S.Departement of Agriculture (https://fdc.nal.usda.gov/). Below are the energy conversion factors:

```
##       banana      cabbage      cassava common_beans       cowpea    groundnut
##          890          250         1600         1480         3360         5500
## irish_potato        maize        mango        onion       pepper    pigeonpea
##          580         3650          800          720          200          810
##     rapeseed         rice      sorghum      soybean    sunflower       tomato
##         1270         3600         3390         1470         5700          210
```

GIS information

Based on the GPS coordinates of centroid of the district (or EPA) of the households, we extracted:

- the Dixon farming system classification for Sub-Saharan Africa Dixon et al. 2021

- the Koeppen's Climate Classification from:
  Beck, H.E., et al. (2018) "Present and future Köppen-Geiger climate classification maps at 1-km resolution", *Nature Scientific Data*, 5, 180214 DOI 10.1038/sdata.2018.214

The population density, and the travel time to cities could not be estimated based on these aggregated coordinates.

Summary

We kept only *rural* households with at least one crop or one livestock. In total, the dataset is made of 1582 households, with information on 6340 crop cultivated and 1622 livestock species. The dataset made of the four tables, together with tlu and energy conversion factors are binded together into a `farmhousehold` object and saved into the file `HHDB_SIMLESA_MWI2013.rds`.