Understanding Farm Size Heterogeneity

# Intro

## Why Investigate Farm Size Heterogeneity

**Farm size can tell us a lot about farms**: Discussions on how much food we produce, who is producing it, and how, are becoming more relevant today than ever. In the next 40 years, we will need to produce significantly more food whilst operating within planetary boundaries (Keating et al., 2014). Farm size has become important in current discussions on food-security, development, and sustainability (Meyfroidt, 2017). We know that: farms of different sizes produce different types of food; small farms are more biodiverse; farm-size is important in determining the amount food a farm can produce (Keating et al., 2014; Ricciardi et al., 2021). Farm size is something that is relatively simple to measure, through surveys or plot-boundary mapping, but it can tell us a lot about things which are more difficult to determine.

**Farm size maps could be useful for decision making:** In recent years, there has been increased emphasis on tailoring development interventions (IFAD, 2019). To design tailored interventions that help us sustainably meet our nutritional needs and improve the quality of lives of farming households, we need to know more about the types of farms which exist, and where we can find them. This is difficult, farmers are highly heterogeneous, there is no fixed definition of “farm type”, and there is not sufficient data to classify farms on a global scale. Instead, researchers have tried to leverage the information associated with farm-size. For example, Herrero *et al.* (2017) used subnational maps on farm size to estimate the contributions different farms make to global nutrient production.

## The Problem

**The major farm size map fails to account for local heterogeneity:** The data they used was produced by Lowder *et al.* (2016). The Lowder map has been widely cited in international agricultural research. The method used to produce the map method scales well, but does not account for skewness in the farm size distributions. Like many indicators of asset ownership, farm size could be positively skewed. One report suggests that in sub-Saharan Africa, 89.6% of farms are classed as “small farms”, they operate on 14.7% of the agricultural land, and that many farm size is being driven downwards (i.e. larger farms controlling more of the land) (GRAIN, 2014). If this is true, the Lowder method could vastly underestimate the number of “extremely small farms”.

**The data collection environment makes it difficult to account for local heterogeneity:** Understanding farm-size distributions, and how they vary between locations is a difficult task. It requires sufficient density of data, at the subnational level to fit a distribution, and it requires data with large enough coverage to understand changes in distributions over larger scales. These data are rare in agricultural research for development. For example, Lowder *et al.* (2016)compared their estimates with data from national agricultural censuses to verify their methods. While these datasets are representative at the national level, they often do not contain sufficient data at the subnational level to capture local variation, sometimes containing as few as 8 data points in some subnational units.

# Objectives

We aim to understand variations in farm-size distributions between different subnational areas. We use a harmonised set of household surveys, covering 276 subnational regions across West, East, and Central Africa.

**Assess/describe farm-size distribution by location:** We begin with an initial assessment to determine whether non-normal distributions of farm-size are common. We investigate characteristics of these farm-size distributions and discuss what the implications of these are in relation to the over/underestimation of the numbers of small farms.

**Question 1:** How useful is it to map mean farm size, if we are interested in counting the smallest farms?

**Model/Predict Farm Size Distributions:** Here we try to see if we can predict farm-size distributions across different subnational locations. Predicting conditional We compare the effectiveness of the four main types of distributional regression model (Kneib et al., 2021): Generalised Additive Models for Location, Scale, and Shape (GAMLSS); Conditional Transformation Models and Distribution Regression; Density Regression; Quantile/Expectile Regression. We assess the models in terms of accuracy and interpretability.

**Question 2:** What are the “best” methods for predicting farm size distributions across locations?

# Proposed Work

Diagram

Description automatically generated

Figure . Basic schematic of analysis plan. First step is data cleaning, mapping, and determining whether there are non-normal distributions of farm-size across locations. Second is the iterative modelling, here we identify drivers, link to data, build models, evaluate and refit. At the third stage, we compare the performance of our different modelling methods, seeing how well they perform in different locations. Finally, we have our “future work” stage. This is how we inte

# Work to Date

## Data Preparation

**Data Prep and Cleaning Ongoing:** I prepared the RHoMIS dataset using the RHoMIS-R-package. I am due to prepare a new iteration in the coming week.

## Data Linkage

**RHoMIS Linked to Census and Satellite Data:** I wanted to link RHoMIS to 2 types of data. The first was census data (international census data provided by IPUMS), and the second was satellite/satellite derived indicators. I identified 9 countries where we have harmonized census data, satellite data and RHoMIS surveys. Census statistics were provided for subnational regions. I linked RHoMIS dataI linked summary statistics from the census to RH. Finally I have developed a procedure for linking RHoMIS to satellite information provided by Google Earth Engine (more on that in the TODO section below).

## Data Coverage

**Country Identification:** The RHoMIS dataset has most surveys in:

* West Africa: Burkina Faso
* East Africa Ethiopia, Kenya, and Tanzania
* Central Africa: Rwanda

However, when we look at how many subnational units contain RHoMIS surveys, we see only Burkina Faso, Rwanda, and Tanzania have RHoMIS surveys in over 50% of their subnational units. Not necessarily a problem, but definitely something to consider.

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

The number of surveys in each subnational unit varies significantly. For example, in Tanzania there is wide spatial coverage, but there are only a few surveys in each unit. This contrasts with Ethiopia, where the number of surveys per unit is much higher, but the .

Chart, bar chart

Description automatically generated

Not sure yet whether it is better to focus on specific countries or identify the subnational units with sufficient data to capture a “distribution” in farm-size.

## Distributions

Next step in the analysis was to look at whether there were patterns in variation of farm size, and whether there were indications of non-normal distributions. The graph below (left) shows the mean land cultivated, against the coefficient of variation. We can see that for locations with a lower mean farm-size, there are regions where the coefficient of variation is very large, and where the coefficient of variation is vary small. On the right graph, we can see that most locations in the dataset have a positive skew, suggesting that maybe modelling mean farm size by location is not helpful for counting the number of small farms.

Chart, scatter chart

Description automatically generatedChart, histogram

Description automatically generated

Finally, the graph below is trying to show the damage that could be done by using the mean to estimate the number of farms of a particular size. Let’s saw we are interested in the proportion of households with a farm-size of less than 2ha. This graph shows the proportion of households with less than 2ha of land, plotted against the mean land cultivated. The black range bars represent the maximum and minimum proportions for 5 farm-size quantiles (q1, q2, q3, q4, q5).

For locations in the lowest land quantile (q1 ≈ 0.64ha), some areas have 100% of households with land less than 2ha, other areas have 75% of households with land less than 2ha.

Now lets take areas in the middle land quantile (q3 ≈ 2.23ha). Regions in this quartile have up to 83% of farms with less than 2ha, down to 0% of farm with less than 2ha of land. In areas like this, mean farm size does not fully explain variation (or the number of different farms of different sizes.

Chart, histogram

Description automatically generated

# Potential Covariates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Description | Reason | Data Type | Source |
| Elevation |  |  |  | [CGIAR/NASA/GEE](https://developers.google.com/earth-engine/datasets/catalog/CGIAR_SRTM90_V4) |
| Population density |  |  |  | [GPWv411/GEE](https://developers.google.com/earth-engine/datasets/catalog/CIESIN_GPWv411_GPW_Population_Density) |
| Data Context |  |  |  | [GPWv411/GEE](https://developers.google.com/earth-engine/datasets/catalog/CIESIN_GPWv411_GPW_Data_Context#bands) |
| Land Cover Class |  |  |  | [Copernicus](https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_Landcover_100m_Proba-V-C3_Global)/Landsat |
| Topographic Diversity |  |  |  | [GEE/ALOS](Global%20ALOS%20Topographic%20Diversit) |
| Global Administrative Layers (layer1) |  |  |  | [FAO/GEE](https://developers.google.com/earth-engine/datasets/catalog/FAO_GAUL_2015_level1) |
| Global Administrative Layers (layer2) |  |  |  | [FAO/GEE](https://developers.google.com/earth-engine/datasets/catalog/FAO_GAUL_2015_level2) |
| Net Primary Production |  |  |  | [FAO/GEE](https://developers.google.com/earth-engine/datasets/catalog/FAO_WAPOR_2_L1_NPP_D#description) |
| Reference Evapotranspiration |  |  |  | [FAO/GEE](https://developers.google.com/earth-engine/datasets/catalog/FAO_WAPOR_2_L1_RET_D) |
| Land/Water in a region |  |  |  | [NASA/GEE](https://developers.google.com/earth-engine/datasets/catalog/GLCF_GLS_WATER#bands) |
| SPEI | Standardized Precipitation Evapotranspiration Index (SPEI) where climatic water balance was aggregated for the last 5 years |  |  | [GRIDMET/GEE](https://developers.google.com/earth-engine/datasets/catalog/GRIDMET_DROUGHT#bands) |
| NDVI |  |  |  | [NASA/USGS/GEE](https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD13A2) |
| Monthly Precipitation |  |  |  | [NASA/GEE](https://developers.google.com/earth-engine/datasets/catalog/NASA_GPM_L3_IMERG_MONTHLY_V06) |
| Soil Moisture |  |  |  | [NASA/GSFC](https://developers.google.com/earth-engine/datasets/catalog/NASA_USDA_HSL_SMAP10KM_soil_moisture)/GEE |
| Nightime lights |  |  |  | [EOG/GEE](https://developers.google.com/earth-engine/datasets/catalog/NOAA_DMSP-OLS_CALIBRATED_LIGHTS_V4) |
| Corrected nighttime lights |  |  |  | [EOG/GEE](https://developers.google.com/earth-engine/datasets/catalog/NOAA_VIIRS_DNB_MONTHLY_V1_VCMSLCFG) |
| Daytime land surface temp |  |  |  | [OMAP/GEE](https://developers.google.com/earth-engine/datasets/catalog/Oxford_MAP_LST_Day_5km_Monthly#description) |
| Travel time to healthcare |  |  |  | [MAP/GEE](https://developers.google.com/earth-engine/datasets/catalog/Oxford_MAP_accessibility_to_healthcare_2019) |
| Gross Primary Production |  |  |  | [NTSG/GEE](https://developers.google.com/earth-engine/datasets/catalog/UMT_NTSG_v2_MODIS_GPP) |
| Cropland |  |  |  | [GFSAD30/GEE](https://developers.google.com/earth-engine/datasets/catalog/USGS_GFSAD1000_V1) |
| **World Clim Bio** |  |  |  | [Berkely/GEE](https://developers.google.com/earth-engine/datasets/catalog/WORLDCLIM_V1_BIO#bands) |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

# References

GRAIN, 2014. Hungry for land: small farmers feed the world with less than a quarter of all farmland [WWW Document]. URL https://grain.org/article/entries/4929-hungry-for-land-small-farmers-feed-the-world-with-less-than-a-quarter-of-all-farmland (accessed 9.7.22).

Herrero, M., Thornton, P.K., Power, B., Bogard, J.R., Remans, R., Fritz, S., Gerber, J.S., Nelson, G., See, L., Waha, K., Watson, R.A., West, P.C., Samberg, L.H., van de Steeg, J., Stephenson, E., van Wijk, M., Havlík, P., 2017. Farming and the geography of nutrient production for human use: a transdisciplinary analysis. Lancet Planet. Health 1, e33–e42. https://doi.org/10.1016/S2542-5196(17)30007-4

IFAD, 2019. Revised Operational Guidelines on Targeting. IFAD.

Keating, B.A., Herrero, M., Carberry, P.S., Gardner, J., Cole, M.B., 2014. Food wedges: Framing the global food demand and supply challenge towards 2050. Glob. Food Secur., SI: GFS Conference 2013 3, 125–132. https://doi.org/10.1016/j.gfs.2014.08.004

Kneib, T., Silbersdorff, A., Säfken, B., 2021. Rage Against the Mean – A Review of Distributional Regression Approaches. Econom. Stat. https://doi.org/10.1016/j.ecosta.2021.07.006

Lowder, S.K., Skoet, J., Raney, T., 2016. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. World Dev. 87, 16–29. https://doi.org/10.1016/j.worlddev.2015.10.041

Meyfroidt, P., 2017. Mapping farm size globally: benchmarking the smallholders debate. Environ. Res. Lett. 12, 031002. https://doi.org/10.1088/1748-9326/aa5ef6

Ricciardi, V., Mehrabi, Z., Wittman, H., James, D., Ramankutty, N., 2021. Higher yields and more biodiversity on smaller farms. Nat. Sustain. 4, 651–657. https://doi.org/10.1038/s41893-021-00699-2

# Notes and Thoughts (for Léo)

K-L divergence is a measure of the information gained, by using one distribution over another.