TAMASA action plan

# What has been done

## Rewrite and document the data cleaning workflow:

Some operations on the data had not been implemented in R code and some key assumptions made, were not explicitly transcribed. It was necessary to make the workflow more transparent. The whole data cleaning workflow is now clearly described and illustrated in one html document. We now have a better idea of the content and limitations of the TAMASA APS17 dataset.

### Recap variables needed to run Paul’s analysis.

Below is a table summarizing all variables needed to run Paul’s analysis. We indicated whether or not variables can be derived from the TAMASA data.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | From TAMASA | comments | From Expert |
| price local maize seed (Tsh/kg) | 1 |  | 0 |
| price hybrid maize seed (Tsh/kg) | 1 |  | 0 |
| price local pigeon pea seeds (Tsh/kg) | 0 |  | 1 |
| price improved pigeon pea seeds (Tsh/kg) | 0 |  | 1 |
| price urea (Tsh/kg) | 1 |  | 0 |
| price CAN (Tsh/kg) | 1 |  | 0 |
| price DAP (Tsh/kg) | 1 |  | 0 |
| price N (Tsh/kg) | 1 |  | 0 |
| price P (Tsh/kg) | 1 |  | 0 |
| price NPK (Tsh/kg) | 1 |  | 0 |
| price TSP (Tsh/kg) | 1 |  | 0 |
| price insecticide (Tsh/kg) | 1 | refer to as pesticide | 0 |
| price fungicide (Tsh/L) | 1 |  | 0 |
| price herbicide (Tsh/L) | 1 |  | 0 |
| selling price maize (Tsh/kg) | 1 |  | 0 |
| selling price pigeon pea (Tsh/kg) | 0 |  | 1 |
| land preparation (day/ha) | 1 |  | 0 |
| sowing (day/ha) | 1 | refer to as planting | 0 |
| fertilizer application (day/ha) | 1 |  | 0 |
| herbicide application (day/ha) | 0 |  | 1 |
| weeding (day/ha) | 1 |  | 0 |
| harvesting maize (day/ton) | 1 |  | 0 |
| harvesting legume (day/ton) | 1 |  | 0 |
| harvesting (day/ton) | 1 |  | 0 |
| treshing maize (day/ton) | 0 |  | 1 |
| treshing legume (day/ton) | 0 |  | 1 |
| treshing (day/ha) | 0 |  | 1 |
| TOTAL | 20 |  | 7 |

Problematic variables are highlighted in red.

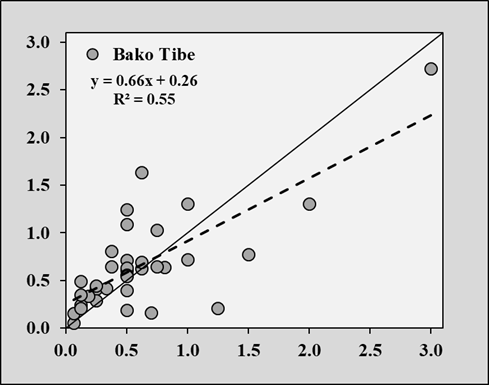
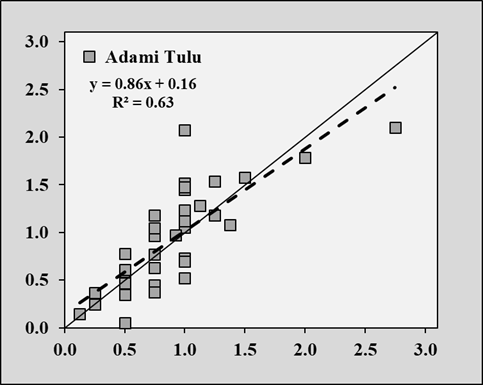
### TAMASA data quality

The data required to rerun Paul’s analysis falls into two broad categories: prices and labour requirements.

Price data seems to be the most reliable. Values were in line with that was reported by Paul and Esther Mungi who is doing her field work in Tanzania.

Unfortunately, labour data does not seem to be as trustworthy as the price data. Mainly because it is impacted by several kinds of uncertainties.

1. Labour variables are calculated based on farmers estimate for:
   1. **Plot area**: We would expect some bias here. Certain plots were supposed to have their area measured by GPS but the data is simply not there. Within TAMASA measurement accuracy check is not possible. However, Marloes did compare farmers’ estimate to GPS data in Ethiopia. See figures below.



*x-axis: area estimated by farmers / y-axis: area measured.*

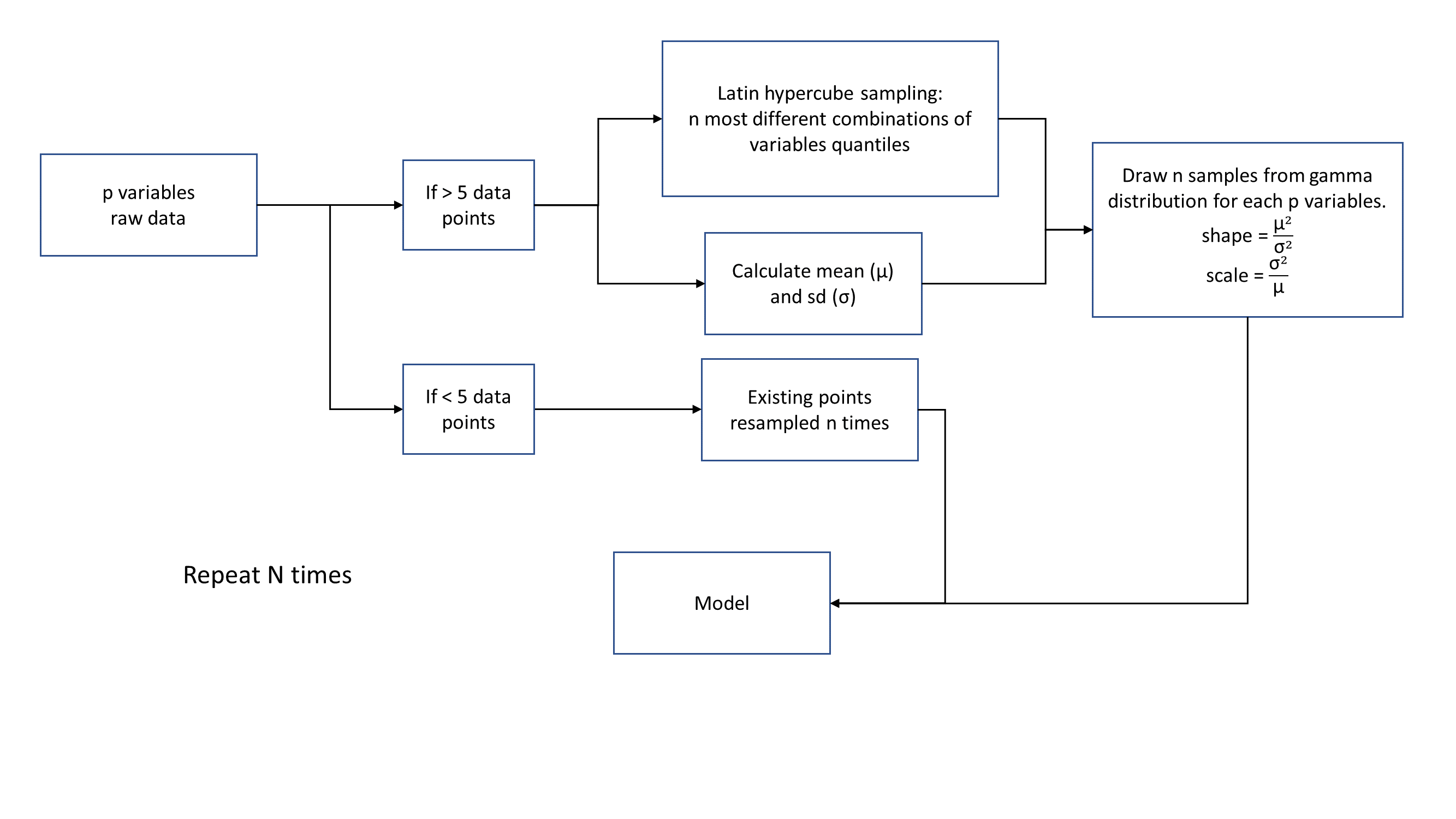
* 1. **Yield**: This variable seems particularly error prone since most farmer in our study area were surveyed before harvest.

1. The wide spread of the observed values suggest a flawed interpretation of the survey questions by the farmers. Those variables were expected to reflect the time spent in the field for a given task in days. However, we do think it might have been interpreted by the farmer as the time period in days over which a given task was performed.

*Comparison literature might come here*

## Implement Sampling Scheme

R code was written to implement a Latin hypercube resampling scheme similar to what was done by van Loon, et al. in their 2019 Global Change Biology paper. This will allow to perform sensitivity analysis in a sounder way. Here below, a schematic of the resampling process.



# What’s next

## Find solution regarding labour data

1. Collect new data on labour requirement for cropping system of interest in Northern Tanzania. *Mukoma....*
2. Drop labour component in trade-off analysis. Keep only productivity versus GHG emissions.

## Re-write simulation code

1. Reformat Paul’s code to make it understandable and reusable.
2. Couple it with resampling scheme.
3. Do test run with existing and/or simulated data.

The main objective is to create a well-documented set of tools ready to be used by anyone. A safe estimate of the amount of work needed would be at least 1 month full time.