Use of Small area estimation for rare crops data in agricultural statistics: case of Burkina Faso and Brazil

Dramane Bako*
International consultant in agricultural statistics
Global Strategy to Improve Agriculture and Rural Statistics
FAO, Statistics Division
Dramane.Bako@fao.org

Flavio Bolliger
Research Coordinator
Global Strategy to Improve Agriculture and Rural Statistics
FAO, Statistics Division
Flavio.Bolliger@fao.org

Abstract

In many countries, agricultural statistics cover many crops of various categories (cereals, horticulture, root crops etc.) A number of crops may be rare considering the number of holdings who produce them (egg less than 10%) and some of them may have high economical values. Thus countries have a great interest on these crops and request national level data for decision making in order to support their production. The estimation of data on rare items can be a serious issue for agricultural statisticians because only a few number of these items appear in the periodical sample agricultural surveys. This paper will discuss data collection issues and the use of small area estimation methods for rare crops using data from Burkina Faso and Brazil.

Keywords: Agricultural statistics; Small area estimation.

1. Introduction

Agriculture is an important economic sector in many developing countries. For instance, in Burkina Faso, the agriculture sector contributed to 35.3 percent of Gross National Product and accounted for more than 37 percent of national exports in 2009 (MAFAP, 2013). For these countries, the sector is then of high importance for food security, poverty reduction and economic growth.

The development of agricultural value chains is among key priorities of most developing countries. However the major crops produced by farmers are not always those, which have high values added. Considering both in-country and international demand and prices, some crops produced by a small number of farmers can have high gross margins for individual farmers and may better contribute to the global contribution of the agricultural sector to the national economic growth. Thus it may appear important to countries to develop relevant policies to promotion the production of these crops and that lead to a great demand of reliable data on these rare crops.

Unfortunately, the usual agricultural surveys generally do not cover sufficient farmers producing these kinds of crops for direct estimates of agricultural variables of interest. The aim of this paper is to explore the use of small area estimation methods for rare crops statistics with applications in Burkina Faso and Brazil. Because of time and data accessibility constraints, we will present in this version of

the document only an application of small area estimation for the estimation the area of fonio in Burkina Faso.

The fonio (***Digitaria exilis***) is a cereal crop produced in Burkina Faso and many other african countries. The demand of this crop is quite high and it is easier to cultivate because it is drought resistant and adapted to the climat of most regions in Burkina Faso. In addition, it has a great potential for food security (higher yield than other popular cereal) and nutrition. It is increasingly highly prized by nutritionists for its richness in methionine and cysteine, two amino acids vital to human health but deficient major cereals such as wheat, rice, maize, sorghum or barley (Vodouhe and Thierno, undated). Since 1974, Fonio is in the the USA Academy of Sciences priority list for underexploited African tropical plants with promising economic value in West and Central Africa (Vodouhe and Thierno, undated). Data on this crop is thus of great important for both the Government of Burkina Faso and its international partners for policy planning.

2. Section 2 : presentation of the data and preliminary issues

For this work, we used the data from two main sources:

i. **the general census of agriculture of 2007 (GCA 2007)**: this census allowed a complete enumeration of all the agricultural holdings of the country
ii. **the annual agricultural survey of 2008**: The main domain of this survey was the province which an administrative entity for which reliable data are required. The total size of sample was 6890 agricultural holdings. This survey collected detailed data on the two major crops planted in each parcel of the sampled holdings. The sample has been selected from the frame developed with the GCA 2007 data.

From these two sources, we can notice from the GCA 2007 that only 6.09% of holdings produced fonio in 2007 and only 2.47% of the sampled holdings in the survey of 2008 were producing fonio (see table1 below). Thus one issue with rare crops appears here: the under coverage of these items in the sample because of their low probability of selection linked to their low number in the population. However this issue can be well addressed through a stratification in the sampling design if the statistical units producing these crops can be identified in the frame before sampling.

Table1: number of holdings from GCA2007 and 2008 survey.

	Census 2007	Survey 2008
Holdings	1 219 241	6 890
Holdings producing fonio	74 239	170
%	6.09%	2.47%

Another issue is the bad repartition of the holdings producing the rare crops in the sample in the main domains. In the table2 below, we can notice that none of the producers selected in the sample in many provinces (main domains) produces fonio although there are many producers of fonio in these provinces. There are only 10 provinces for which the sample selected included some producers of fonio. Here again, this issue can be well addressed through a good allocation of the sample if the statistical units producing these crops can be identified in the frame before sampling.

3. Methodology and results

Our main objective is to produce reliable estimate of the total cultivated fonio area in the main domains and the average area per holding. Thus using the average yield of fonio from survey or secondary data, we will be able to estimate the production of fonio in the country.

3.1. Analysis of direct estimates

In order to assess the reliability of direct estimates of fonio data using the 2008 survey data, we have calculated the total number of holdings producing fonio in 2008 using the weight of the sample and the average area planted for fonio in the 45 provinces (main domains) and at the national level. Two major observations can be made here from table2 below:

1. The total number of holdings producing fonio in the country in 2008 using the survey data is 33 356 which represents a decrease of 55% compared to the previous year figure (74 239) from the census. Such a decrease is obviously inconsistent and reveals a lack of reliability of the direct estimates.
2. The coefficient of variation of the total planted area for fonio is very high in most main domains and at the national level (105%).

Table2: direct estimations results

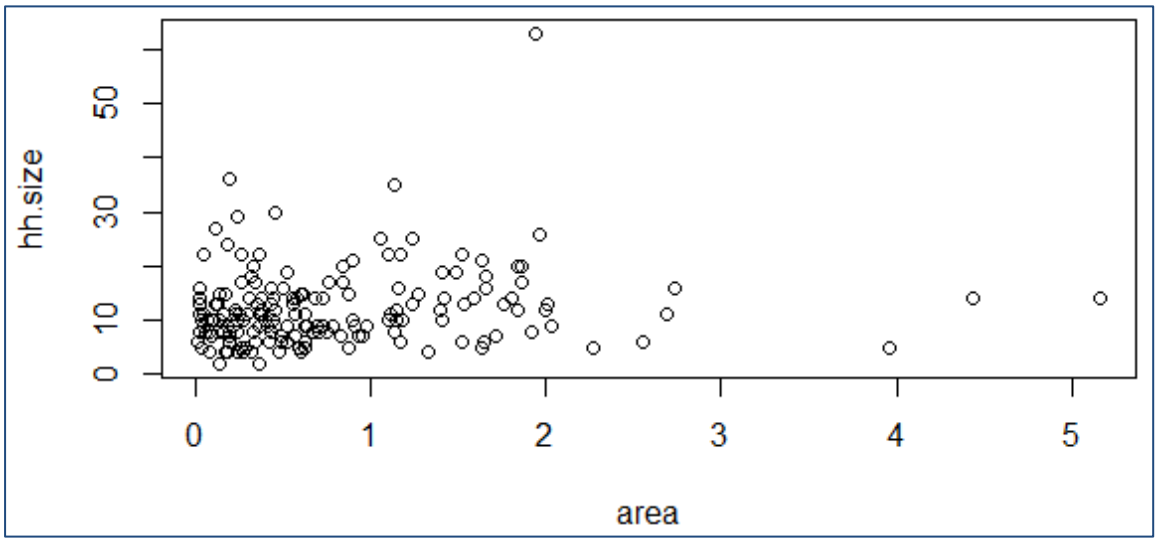
Provinces	Planted area (2008 survey)		Producers of fonio from the 2008 sample survey		Producers of fonio from the 2007 census
	Mean	Coefficient of variation	unweighted	weighted	
BAM					992
BAZEGA					196
BOUGOURIBA					152
BOULGOU					1 159
BOULKIEMDE					454
COMOE	2,75	58,5%	5	766	1 842
GANZOURGOU					395
GNAGNA					347
GOURMA					438
HOUET	0,83	74,1%	27	3 084	4 391
KADIOGO					269
KENEDOUGOU	1,19	69,5%	29	3 563	3 092
KOSSI	0,77	89,8%	64	17 092	21 336
KOURITENGA					260
MOUHOUN					759
NAHOURI					497
NAMENTENGA					388
OUBRITENGA					263
OUDALAN					210
PASSORE					286
PONI					271
SANGUIE					325
SANMATENGA					752

Provinces	Planted area (2008 survey)		Producers of fonio from the 2008 sample survey		Producers of fonio from the 2007 census
	Mean	Coefficient of variation	unweighted	weighted	
SENO					284
SISSILI					342
SOUM	0,03	10,9%	2	1 232	4 843
SOUROU					624
TAPOA	0,02		1	247	1 202
YATENGA	0,12	117,5%	9	3 059	12 517
ZOUNDWEOGO					312
BALE					369
BANWA	0,35	71,5%	3	1 595	2 547
IOBA					774
KOMANDJOARI					32
KOMPIENGA					248
KOULPELOGO					703
KOURWEOGO					107
LERABA	0,54	73,2%	21	1 930	2 251
LOROUM	0,13	85,4%	9	789	6 518
NAYALA					402
NOUMBIEL					47
TUY					393
YAGHA					94
ZIRO					195
ZONDOMA					361
Total	0,79	101,2%	170	33 356	74 239

3.2. Small area estimation

Considering the unreliability of direct estimations, we explored the use of a small area estimation model to improve the estimates. The small areas here are the farmers producing fonio in the main domains of the 2008 sample survey. Considering the nature of our data, the GREG model appears suitable for the estimation of the total area planted. We followed the following steps in our approach:

Step1: Identification of variables correlated to the planted area and collected during both the 2007 census and the 2008 survey.
Using the 2008 survey data, we find that only the size of the household is correlated to the planted area and the correlation is significant at a 0.05 level



Step2: regression
We performed a fixed effect regression with the planted area as dependent variable and the size of the household as independent variable using the 2008 survey data

Step3: estimation
Following Rao (2003) quoted by Global Strategy (2015), the estimate of average area in each small area i using the GREG estimator ($\hat{\bar{Y}}_i^{GREG}$) is:

$$\hat{\bar{Y}}_i^{GREG} = \frac{1}{N_i}\sum_{j\in S_i} y_{ij} + \frac{1}{N_i}\left(\sum_{j\in U_i} X_{ij}^T - \sum_{j\in S_i} X_{ij}^T\right)\hat{\beta}$$

Where $\hat{\beta} = (\sum X_{ij}X_{ij}^T)^{-1}(\sum X_{ij}y_{ij})$

N.B. For steps 2 and 3, we used the R package JoSAE following the instructions of Global Strategy (2015).
In case there is no unit sampled in a given small area, the average planted area is estimated using the following GREG estimator:

$$\hat{\bar{Y}}_i^{GREG-OUT} = \frac{1}{N_i}\sum_{j\in U_i} X_{ij}^T\,\hat{\beta}$$

The total planted area in each domain is then estimated by multiplying the average area planted estimated with the GREG model by the total number of holdings producing fonio in the domain.

3.3. Results
For the 10 provinces for which the sample selected included producers of fonio, the model provides the small-area GREG estimates of the average planted area and the MSE (table3).
Table3: SAE results

Provinces	GREG est	GREG mse
COMOE	2,77	0,70
HOUET	0,80	0,12
KENEDOUGOU	1,19	0,15

KOSSI	0,69	0,08
SOUM	-0,02	0,03
TAPOA	0,10	NA
YATENGA	0,03	0,05
BANWA	0,37	0,14
LERABA	0,50	0,08
LOROUM	0,06	0,03

For the province Soum, the GREG estimation of the average planted area is negative (not possible); we re-estimated this value using the $\hat{Y}_i^{GREG-OUT}$ formula provided above.
The final results are provided in the table4 below:

domain.ID	domain.name	Planted area of fonio (ha)	
		Mean	Sum
1	BAM	0,69	684,75
2	BAZEGA	0,68	134,15
3	BOUGOURIBA	0,67	101,79
4	BOULGOU	0,67	779,66
5	BOULKIEMDE	0,73	333,49
6	COMOE	2,77	5 098,36
7	GANZOURGOU	0,76	299,81
8	GNAGNA	0,64	223,38
9	GOURMA	0,71	310,92
10	HOUET	0,80	3 502,54
11	KADIOGO	0,62	166,60
12	KENEDOUGOU	1,19	3 690,81
13	KOSSI	0,69	14 768,72
14	KOURITENGA	0,77	200,74
15	MOUHOUN	0,67	506,27
16	NAHOURI	0,61	303,92
17	NAMENTENGA	0,71	277,00
18	OUBRITENGA	0,67	175,05
19	OUDALAN	0,65	136,06
20	PASSORE	0,69	198,42
21	PONI	0,64	174,58
22	SANGUIE	0,82	266,71
23	SANMATENGA	0,73	551,80
24	SENO	0,67	190,57
25	SISSILI	0,71	242,82
26	SOUM	0,64	3 075,93
27	SOUROU	0,66	410,07
28	TAPOA	0,10	123,26
29	YATENGA	0,03	375,45
30	ZOUNDWEOGO	0,68	211,14
31	BALE	0,67	245,63
32	BANWA	0,37	944,93
33	IOBA	0,66	508,60
34	KOMANDJOARI	0,70	22,28
35	KOMPIENGA	0,66	164,33
36	KOULPELOGO	0,68	478,79
37	KOURWEOGO	0,74	79,63

domain.ID	domain.name	Planted area of fonio (ha)	
		Mean	Sum
38	LERABA	0,50	1 124,83
39	LOROUM	0,06	376,58
40	NAYALA	0,72	290,31
41	NOUMBIEL	0,68	32,16
42	TUY	0,69	269,84
43	YAGHA	0,63	58,97
44	ZIRO	0,70	137,42
45	ZONDOMA	0,78	280,74
Total			42 529,81

References

Global Strategy (2015). Spatial Disaggregation and Small-Area Estimation Methods for Agricultural Surveys: Solutions and Perspectives. FAO, Rome, Italy.

MAFAP (2013). Review of food and agricultural policies in Burkina Faso. MAFAP Country Report Series, FAO, Rome, Italy.

Rao, J.N.K. 2003. Small Area Estimation. New York, Wiley

Vodouhe and Thierno, undated. Promoting fonio production in West and Central Africa through germplasm management and improvement of post harvest technology, http://underutilized-species.org/Documents/PUBLICATIONS/rv_fonio.pdf accessed on 27/03/2017.