Land Area Measurement bias: Evidence from West African countries

GOUNDAN, Anatole and MAGNE DOMGHO, Léa Vicky

International Food Policy Research Institute, International Food Policy Research Institute

November 2016

Land Area Measurement bias: Evidence from West African countries

Anatole Goundan[1] **& Lea Vicky Magne Domgho**

International Food Policy Research Institute, West and Central Africa Office

Abstract

Planted and harvested areas are crucial for agricultural statistics. In developing countries, such statistics are estimated using farmers' reports which are systematically biased. Given the importance of the area size in designing policy and in farmers' wealth, it is essential to empirically assess that bias for the countries in order to inform the potential impact of that issue in different contexts. This paper, therefore, contributes to analyzing farmers' plot size estimation bias in four West African countries (Burkina Faso, Mali, Niger, and Nigeria). The paper also explores the determinants of the bias in land measurement in these countries. Our findings indicate that the bias in land measurement is a serious issue among West African countries and varies between 14% and 171% (in absolute value) of the correct area size. In terms of the determinants of acreage discrepancy, our findings reveal that the respondents' age, education, land acquisition status, plot size, area unit measurement, are influential.

Keywords: Area estimation, land measurement, GPS measure, acreage gap, West Africa.

JEL classification: C81, O12, Q15, Q12

[1] Corresponding author, Email: a.goundan@cgiar.org
Titre 3396, Lot #2
BP 24063 Dakar Almadies
Senegal

1. Introduction

Agriculture is an important sector in developing countries, and billions of people exclusively depend on it for their income and subsistence. Therefore, agriculture is a strategic area of interest for policymakers in terms of food security, poverty alleviation, biosecurity, etc. Usual components of agricultural policies are based on agricultural land and productivity. As stated by Carletto et al. (2015), "Land is a key measure of absolute and relative farmer wealth, a critical input in production, and a key variable for normalizing agricultural input use and output measures." In addition, to compute crop production, agricultural statisticians use harvested area and yield (Keita and Carfagna, 2009). Therefore, a good estimation of plots is essential and may have huge consequences on agricultural inputs and outputs.

In developing countries, planted area estimation generally rely on farmers' self-reporting. However, it has been shown that self-reported area sizes are systematically biased (DeGroote and Traore, 2005; Carletto et al., 2013, 2015a, 2015b, Holden, S. T., & Fisher, 2013). For example, existing literature shows that small-scale plots of land are overestimated, while large-scale plots are under-reported. In sub-Saharan Africa, most of the farming takes place on small plots of land, less than one hectare in size. Therefore, report bias is crucial and needs to be assessed in order either to educate farmers in estimating their acreage or to systematically measure plots size using

GPS during data collection process as suggested by some authors (Palmegiani, 2009; Keita and Carfagna, 2009; Schoning et al., 2005).

In our days, many authors are interested in explaining the gap between GPS measure of land area and farmers reported sizes. Several factors had been advanced. Among them, respondent or household head characteristics (gender, age, and level of education), plot characteristics (plot slope, plot relief, plot size, canopy cover, etc.), land ownership, weather conditions, or locations (Keita and Carfagna, 2009, Carletto et al., 2013, 2015b). Carletto et al. (2013), using Uganda data, found that larger farms tend to under-report, while small farms over-report. They also showed that older household's heads are less accurate in their reporting of the plot size. Land ownership is shown to have a significant impact on the discrepancy. Carletto et al. (2015b) studied the determinants of acreage gap in the context of three African countries: Ethiopia, Nigeria, and Tanzania. They found that for these different countries, plot size and distance from the plot to the dwelling place have a significant influence on the estimation bias. They also showed that tree cover has little effect.

In the line with this literature, this paper contributes to the debate by assessing the acreage gap between farmers' self-reported area estimation and GPS measure in West African countries. Particularly, this paper will focus on four countries of the region, namely Burkina Faso, Mali, Niger, and Nigeria. In addition, the determinants of acreage discrepancy will be analyzed for each of these countries.

The rest of this paper is organized as follows. The next section provides a description of the data and, especially, the discrepancy observed between self-reported area size and GPS measure.

Section 3 succinctly presents the econometric models used and discussed findings. The last section summarizes the main findings of the analysis and some policy implications.

2. How wide farmer's acreage estimation diverges from GPS measure?

In developing countries, a big share of farmers never attend school. Therefore, it is a big challenge for them to accurately estimate their total area. They do have an idea of how big their farms are and even have some traditional means to measure their planted area. The question therefore is how those traditional (local) knowledge match with advanced land measurement tools such as Global Position System (GPS). No one could answer this question unless they have farmers' declaration and GPS measurement of the same land. Fortunately, recent surveys, conducted by the World Bank team in collaboration with national statistics offices of some developing countries, could be used to shed light on how well local knowledge is in line with modern tools in terms of land measurement. As cases of study, we use data from Livings Standards Measurement Study-Integrated Surveys of Agriculture (LSMS-ISA) for Burkina Faso (2014/2015), Mali (2014/2015), Niger (2014/2015) and Nigeria (2012/2013).

2.1. Case of Burkina Faso

The survey for Burkina Faso, implemented between 2014 and 2015, deals with a total of 29,143 plots belonging to 10,799 farm households. However, only 3,894 plots of land for 1,072 farm households had been measured by GPS. The main reason for this fact is that GPS did not work during data collection. Table 1 reports key summary statistics (mean, median and standard deviation in hectare) for the plots for which information on both self-reported and GPS-measured area measurements are available.

Table 1: Acreage summary statistics at region level in Burkina Faso

	Self-reported area			**GPS-measured area**		
	Average (ha)	Median (ha)	Std. dev. (ha)	Average (ha)	Median (ha)	Std. dev. (ha)
Hauts Bassins	1.6	1	1.34	1.3	0.96	1.25
Boucle du Mouhoun	1.8	1	2.16	1.5	0.99	2.65
Sud Ouest	0.86	0.5	0.73	0.84	0.57	0.96
Centre Ouest	0.73	0.5	0.63	0.88	0.43	3.05
Plateau central	1	1	0.85	0.88	0.61	0.88
Nord	0.93	0.75	0.83	0.67	0.39	0.80
Centre	0.42	0.25	0.34	0.37	0.25	0.35
Cascade	2	1	2.05	1.7	1.1	2.25
Centre sud	0.81	0.5	1.92	0.49	0.3	0.54
Burkina Faso: overall	**1.1**	**0.75**	**1.42**	**1**	**0.6**	**1.94**

Source: Authors' calculation, Burkina Faso LSMS-ISA 2015/2015

At Burkina Faso level, the two methods show quite similar results, based on usual statistics (mean, median and standard deviation). Indeed, the average size of plots using GPS is 1 ha, while the self-reported average is 1.1 ha. The standard deviation of the GPS-measured area is 1.94 ha, a mere 0.52 ha larger than the area reported by farmers. Disaggregating the data by region confirmed that self-reported area is quite larger than that using GPS for all regions except *Centre-Ouest*.

In terms of discrepancy between GPS-measured area (considered as reference) and self-reported area, Table 2 depicts the distribution (mean, median, relative discrepancy, standard deviation, and coefficient of variation) of acreage gap at regional and national level in Burkina Faso. In general, farmers overestimate (negative discrepancy) the size of their plots by on average 0.14 ha (equivalent to 14%). At regional level, exception to the region of *Centre-Ouest* where farmers underestimate area size (17.05%), the area size is overestimated in all regions by a relative gap which varies between 65.31% in *Centre-Sud* and 2.14% in *Sud-Ouest*. The coefficients of variation (CV) at national and regional levels are very large (between 300% and 4,000%) which means that farmers-reported area sizes are very volatile around the "true area size" (GPS-measured area size).

Table 2: Acreage gap summary statistics at region level in Burkina Faso

	Acreage gap (GPS-SR)				
	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Hauts Bassins	-0.31	-0.11	-23.85	1.8	580.65
Boucle du Mouhoun	-0.28	-0.18	-18.67	2.5	892.86
Sud Ouest	-0.018	-0.06	-2.14	0.78	4333.33
Centre Ouest	0.15	-0.08	17.05	3	2000
Plateau central	-0.15	-0.13	-17.05	0.81	540
Nord	-0.27	-0.3	-40.30	0.86	318.52
Centre	-0.047	-0.07	-12.70	0.34	723.40
Cascade	-0.28	-0.19	-16.47	1.7	607.14
Centre sud	-0.32	-0.11	-65.31	1.9	593.75
Burkina Faso: overall	**-0.14**	**-0.13**	**-14**	**1.8**	**1285.71**

Note: CV stands for Coefficient of Variation

Source: Authors' calculation, Burkina Faso LSMS-ISA 2015/2015

In order to check whether small-scale plots are overestimated compared to large-scales plots, we split plots into three groups using the GPS-measured area size: small (<1 ha), medium (>=1 ha and <3 ha), and large (>=3 ha). Table 3 presents the frequency for each farm size, of mean, median, relative gap, standard deviation and CV of area discrepancy. In Burkina Faso, about 69.1% of the plots measure less than 1 ha in size, while only 5.15% of plots are more than 3 ha in size. On average, small plots are overestimated and large plots are underestimated. Small plots are over-reported by 77.5% on average, while large plots are systematically under-reported by on average 31.63%.

Table 3: Acreage gap summary statistics over farm size in Burkina Faso

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Small	69.10	-0.31	-0.18	-77.50	0.79	254.84
Medium	25.75	-0.09	0.09	-5.52	1.01	1122.22
Large	5.15	1.79	1.05	31.63	6.83	381.56

Source: Authors' calculation, Burkina Faso LSMS-ISA 2015/2015

In the literature, authors usually classify farmer's report into three categories: zero discrepancy (good estimation), negative discrepancy (overestimation) or positive discrepancy (underestimation). But it is unrealistic to imagine that farmers were able to have an exact measurement of their plots. Therefore, in this study, we depart a bit from this practice by widening the zero discrepancy group to include all observations with more or less 10% relative discrepancy of GPS measure. This means that all farmers' plots area estimation that do not divert from the GPS measurement of more or less 10% will be considered as a good estimation of their plots size. Consequently, a discrepancy will be themed "underestimated" or "overestimated" when it is more than 10% of the true area size in absolute. Table 4 presents the distribution of discrepancy over the quality of reporting.

Table 4: Acreage gap summary statistics over the quality of report in Burkina Faso

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Good	13.43	0.01	0	0.74	0.1	1000
Underestimation	27.57	0.79	0.37	46.75	2.88	364.56
Overestimation	59	-0.62	-0.34	-105.08	1	161.29

Source: Authors' calculation, Burkina Faso LSMS-ISA 2015/2015

Farmers overestimate plot size in 59% of the cases, underestimate it in 27.57% of plots, and have a fair estimation in only 13.43% of the cases. For the good estimation cases, the relative discrepancy is less than 1%. In the "underestimation" group, the self-reported area is 46.75% less than the GPS measures. Concerning the "overestimation" group, the bias is 105.08% of the real area measure.

2.2. Case of Mali

The LSMS-ISA survey for Mali was implemented between 2014 and 2015. The total number of plots is 9,658 which belong to 2,299 farm households. About 94% of these plots (9,106 plots) have a corresponding GPS measure. Table 5 reports key summary statistics (mean, median and standard deviation in hectare) for the plots for which information on both self-reported and GPS-measured area measurements are available.

Table 5: Acreage summary statistics at region level in Mali

	Self-reported area			**GPS-measured area**		
	Average (ha)	Median (ha)	Std. dev. (ha)	Average (ha)	Median (ha)	Std. dev. (ha)
Kayes	1.20	0.50	4.24	5.63	0.64	15.33
Koulikoro	1.03	0.50	1.54	2.67	0.62	8.98
Sikasso	1.63	0.75	5.16	1.51	0.75	4.31
Segou	1.58	0.86	4.30	3.06	0.77	8.50
Mopti	1.68	1.00	2.95	3.40	1.09	8.71
Tombouctou	1.12	0.50	2.29	4.81	0.75	11.81
Gao	0.91	0.43	1.73	1.43	0.30	5.05
Bamako	0.81	0.01	1.20	0.81	0.01	1.19
Mali: Overall	**1.43**	**0.75**	**3.90**	**2.88**	**0.75**	**8.87**

Source: Authors' calculation, Mali LSMS-ISA 2015/2015

In Mali, unlike Burkina Faso, there is a huge difference between plot size distributions from the two methods. In the overall sample, the average plot size using GPS is 2.88 ha compared to the 1.43 ha estimated by farmers. Farmers underestimate their plots size by half. Results are the same at regional level except for Bamako and Sikasso. GPS-measured area sizes also show a very large variability compared to farmers' reports. The distribution of acreage gap is presented on Table 6.

Table 6: Acreage gap summary statistics at region level in Mali

	Acreage gap (GPS-SR)				
	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Kayes	4.40	0.00	78.22	15	340.91
Koulikoro	1.60	0.01	60.01	8.9	556.25
Sikasso	-0.13	0.00	-8.62	5.4	4153.85
Segou	1.50	0.03	49.03	8	533.33
Mopti	1.70	0.00	50.00	8.9	523.53
Tombouctou	3.70	0.04	76.94	11	297.30
Gao	0.52	0.00	36.34	5	961.54
Bamako	-0.01	0.00	-0.89	0.03	444.44
Mali: Overall	**1.40**	**0.00**	**48.63**	**8.9**	**635.71**

Source: Authors' calculation, Mali LSMS-ISA 2015/2015

In Mali, farmers underestimate largely the size of their plots by on average 1.4 ha (equivalent to 48.63%). Underestimation of land area by almost half has dramatic consequences on agricultural statistics, and especially on acreage, yield, production, and inputs demand and supply. It is then essential for the government to find the best way to change this situation. At the regional level, there is quite a difference in measurement. The regions of Kayes and Tombouctou show up a very large (more than twice the national average) positive discrepancy. Farmers of these regions under-report their plot sizes by more than 3 ha on average. Conversely, in the regions of Sikasso and Bamako, farmers slightly underestimate area sizes respectively by 8.62% and 0.89%. Results for regions reveal that the regions where acreage estimation is made by farmers are more challenging. Those regions are Kayes, Tombouctou Koulikoro, Mopti and Segou where the relative discrepancy is very high, between 49% and 78%. Like for Burkina Faso, there is a big variability of farmers' reported area size around the GPS-based area size. Since plot size may influence the accuracy of farmers' report, Table 7 presents the distribution of acreage gap by plot size groups. As expected, small plots are overestimated and large plots are underestimated. Farmers over-report small plots by about 75% on average, while largest plots are systematically under-reported by on average 74.52%.

Table 7: Acreage gap summary statistics over farm size in Mali

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Small	55.90	-0.3	0	-75.00	3.15	1050.00
Medium	27.46	-0.06	0.07	-3.59	2.06	3433.33
Large	16.65	11.14	3.13	74.52	19.44	174.51

Source: Authors' calculation, Mali LSMS-ISA 2015/2015

In terms of the quality of the estimation, Table 8 shows that 28.26% of plot sizes have approximately a good estimation, 40.87% of underestimation and 30.87% of overestimation. Underestimated areas divert from GPS measurement by 4.52 ha, corresponding to a relative gap of 82.94%. For overestimation cases, on average the gap is at 1.12 ha (149%). Even though the absolute value underestimation cases are higher than underestimated cases, the relative discrepancy is higher for overestimated plots (149.33%) than underestimated plots (82.94%).

Table 8: Acreage gap summary statistics over the quality of report in Mali

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Good	28.26	0.01	0	0.60	0.1	1000.00
Underestimation	40.87	4.52	0.4	82.94	12.76	282.30
Overestimation	30.87	-1.12	-0.3	-149.33	4.93	440.18

Source: Authors' calculation, Mali o LSMS-ISA 2015/2015

2.3. Case of Niger

The LSMS-ISA survey for Niger used in this study was implemented between 2014 and 2015. A total of 5,634 plots corresponding 2,168 farm households were covered. Among these plots, 3,812 plots (67.66%) have both self-reported and GPS measures. Table 9 compares farmers' estimation to GPS measurement of plot size.

Table 9: Acreage summary statistics at region level in Niger

	Self-reported area			GPS-measured area		
	Average (ha)	Median (ha)	Std. dev. (ha)	Average (ha)	Median (ha)	Std. dev. (ha)
Agadez	0.65	0.30	0.99	0.73	0.28	1.15
Diffa	2.06	1.00	2.46	2.47	1.58	3.54
Dosso	2.22	1.50	2.10	1.87	1.25	2.11
Maradi	1.94	1.00	3.18	1.71	1.08	2.29
Tahoua	2.22	1.50	2.37	1.68	1.10	2.30
Tillaberi	3.64	2.00	6.77	3.44	2.08	6.40
Zinder	2.36	1.50	3.16	1.81	1.01	2.72
Niamey	4.96	0.52	17.18	14.24	2.20	24.24
Niger: overall	**2.35**	**1.50**	**3.62**	**1.99**	**1.15**	**3.37**

Source: Authors' calculation, Niger o LSMS-ISA 2015/2015

The overall sample statistics reveal that both methods imply similar results. The average size reported by farmers is 2.35 ha compared to the 1.99 ha produced by GPS measure. This result masks some pervasive heterogeneity detectable at regional level. For example, in the region of Niamey, self-reported average size is three times less than the GPS-measured area size. The summary statistics of acreage gap between the two measurement methods are presented in Table 10.

Table 10: Acreage gap summary statistics at region level in Niger

	Acreage gap (GPS-SR)				
	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Agadez	0.089	-0.019	12.11	0.74	831.46
Diffa	0.41	0.24	16.59	3.2	780.49
Dosso	-0.35	-0.29	-18.69	2.4	685.71
Maradi	-0.22	-0.21	-12.84	3.2	1454.55
Tahoua	-0.54	-0.33	-32.22	2.8	518.52
Tillaberi	-0.2	0	-5.82	9.1	4550.00
Zinder	-0.54	-0.35	-29.78	3.1	574.07
Niamey	9.3	0.52	65.32	17	182.80
Niger: overall	**-0.36**	**-0.24**	**-18.06**	**4.2**	**1166.67**

Source: Authors' calculation, Niger o LSMS-ISA 2015/2015

Plots size is overestimated in Niger by farmers. They over-reported their plot sizes by 18% corresponding to a 0.36 ha gap on average. From one region to another, the quality of farmers' estimation is not similar. In three out the eight regions in Niger, the plot sizes are underestimated. The largest underestimation rate is observed in the region of Niamey where farmers report 9.3 ha (65.32%) less than the real plot size. Besides, the largest overestimation rate is observed in the region of Tahoua (32.22%). Even though the relative discrepancy in different regions of Niger is not as high as for Burkina Faso and Mali, it is sufficiently important to imply actions from the government.

Table 11 presents the distribution of area discrepancy by plot size. As for Burkina Faso and Mali, small plots are overestimated and large plots are underestimated. The reported overestimated area for small plot sizes represents on average 152.73% of the corresponding GPS measure. Medium plot size farmers also over-report their area size by about 44.71% on average, while largest plots are under-reported by 27.78%.

Table 11: Acreage gap summary statistics over farm size in Niger

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Small	40.18	-0.84	-0.48	-152.73	1.6	190.48
Medium	38.54	-0.76	-0.08	-44.71	3.7	486.84
Large	21.28	1.75	1.33	27.78	7.82	446.86

Source: Authors' calculation, Niger o LSMS-ISA 2015/2015

In terms of quality of farmers' estimation (see Table 12), only 13.84% of plots have a fair self-reported area size, 35.18% of underestimation and 50.98% of overestimation. The relative discrepancy is 1.40% for good estimation cases, 54.99% for underestimated plots, and 150% for overestimation cases.

Table 12: Acreage gap summary statistics over the quality of report in Niger

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Good	13.84	0.03	0	1.40	0.23	766.67
Underestimation	35.18	1.93	0.84	54.99	4.63	239.90
Overestimation	50.98	-1.7	-0.81	-150.44	3.94	231.76

Source: Authors' calculation, Niger o LSMS-ISA 2015/2015

2.4. Case of Nigeria

The LSMS-ISA survey for Nigeria used in this study is the second wave survey which was conducted between 2012 and 2013. The total number of plots is 6,054 which belong to about 3,000 farms households. About 87% of these plots (5,279 plots) have two distinct measures of their size (self-reported and GPS). Table 13 reports key summary statistics (mean, median and standard deviation in hectare) for the two measurements of plots size.

Table 13: Acreage summary statistics at region level in Nigeria

	Self-reported area			GPS-measured area		
	Average (ha)	Median (ha)	Std. dev. (ha)	Average (ha)	Median (ha)	Std. dev. (ha)
North Central	1.44	0.288	4.98	0.44	0.27	0.60
North East	1.64	1	3.80	0.78	0.52	0.95
North West	1.37	0.8	3.91	0.51	0.27	0.83
South East	0.74	0.072	5.19	0.10	0.06	0.12
South South	1.52	0.2668	8.87	0.29	0.13	0.84
South West	0.90	0.48	1.28	0.71	0.35	0.91
Nigeria: overall	**1.29**	**0.494**	**4.81**	**0.47**	**0.24**	**0.78**

Source: Authors' calculation, Nigeria LSMS-ISA 2012/2013

In Nigeria, like Mali, there is a huge difference between plot size distributions from the two methods. In the overall sample, the average plot size reported by farmers (1.29 ha) is three times the one measured using GPS (0.47 ha). It seems that farmers, in Nigeria, overestimate largely their plot sizes. Results are the same at the regional level where the average self-reported plot size is always greater than GPS-measured plot size. In addition, Farmers' estimation is more volatile (4.81 ha as standard deviation) than GPS measurement (0.78 ha as standard deviation). Concerning the observed gap between both measures, Table 14 depicts its distribution at the national level and across zones in Nigeria.

Table 14: Acreage gap summary statistics at region level in Nigeria

	Acreage gap (GPS-SR)				
	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
North Central	-1	-0.062	-226.14	4.9	490.00
North East	-0.86	-0.31	-110.88	3.9	453.49
North West	-0.86	-0.5	-168.53	3.9	453.49
South East	-0.64	-0.013	-649.60	5.2	812.50
South South	-1.2	-0.086	-416.02	8.9	741.67
South West	-0.19	-0.063	-26.90	1.3	684.21
Nigeria: overall	**-0.81**	**-0.13**	**-171.38**	**4.8**	**592.59**

Source: Authors' calculation, Nigeria LSMS-ISA 2012/2013

In Nigeria, at the national level and all zones, farmers systematically overestimate the size of their plots. At the national level, the overestimation rate is high and equals 171.38% of the GPS measurement. This means that farmers declare a plot size equivalent to almost three times the correct plot size. Therefore, the direct use of such estimation for statistics agencies may lead to a wrong estimation of acreage, yield, and production. At regional level, the South-East and South-South areas are the ones where the highest overestimation rate are observed. A special attention should be paid to those areas and particular strategies should be explored to help farmers better estimate their plot sizes. The South-West area showed the lowest over-reporting rate (26.90%).

Table 15: Acreage gap summary statistics over farm size in Nigeria

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Small	88.64	-0.9	-0.15	-321.43	4.93	547.78
Medium	9.53	-0.43	0.03	-27.04	2.97	690.70
Large	1.82	1.96	2.17	40.08	4.49	229.08

Source: Authors' calculation, Nigeria LSMS-ISA 2012/2013

Table 15 presents how measurement discrepancy is distributed across plot size groups. The main information is that almost all plots (88.64%) in this survey are less than 1 ha in size. Furthermore, small plots are largely (321.43%) overestimated. As for all other countries, larger plots size are always under-reported (40.08%).

Table 16: Acreage gap summary statistics over groups of declaration in Nigeria

	Frequency (%)	Average (ha)	Median (ha)	Average relative gap (%)	Std dev. (ha)	CV (%)
Good	9.28	0.01	0	1.49	0.07	700.00
Underestimation	23.18	0.47	0.15	55.95	0.97	206.38
Overestimation	67.54	-1.31	-0.36	-396.97	5.64	430.53

Source: Authors' calculation, Nigeria LSMS-ISA 2012/2013

Using our algorithm of classification of measurement discrepancy in different quality groups, Table 16 reveals that only 9.28% of self-reports can be considered as good estimation, while 67.54% of plots are overestimated. Results show that overestimation is a big issue in Nigeria, with a relative average discrepancy of about 400%. In this condition, authorities have to invest on approaches that guaranty a good quality of collected data during surveys.

3. Determinants of acreage discrepancy in West African countries

3.1. The econometric approach

The following section explores the factors affecting the acreage bias between GPS and farmers' estimation. First, as in Carletto et al., (2013 and 2015a) we use a simple econometric method to identify the determinants of acreage bias. The following model will be estimated:

$$Y_i = \beta X_i + \varepsilon_i \qquad (1)$$

Where Y_i is the plot size specific difference between the GPS and the self-reported measure, X_i is a (K + 1)-row vector of control variables with '1' as its first element, $\beta = (\beta_0, \beta_1, \ldots, \beta_K)'$ is vector of parameters to be estimated, ε_i is a two-sided error term representing white noise. Explanatory variables include the respondents' characteristics (age, education, gender, and ethnic group), locations, etc. To control the bias linked to the GPS measure, we include plot characteristics (plot relief, plot size, plot unit area used by farmers), because GPS measure may vary systematically with the plot size, slope, shape, and the presence of trees (Magezi-Apuuli et al., 2005; Keita and Carfagna, 2009; Carletto et al., 2013). We also add plot size and its squared term, to test whether small and large plot size differently affect the dependent variable. Other variables such as land ownership (property rights) and weather conditions are useful and will be included when available.

Model (1) (column 1 of regression results tables) gives an idea of the variables that potentially influence the gap between farmers' declarations and the GPS measurement. But the results of this regression could not be interpreted unequivocally. For example, a variable with a positive coefficient can only be interpreted in relation to the observed value of the gap (underestimation or overestimation). In other words, in a case of underestimation (positive gap in our case), this factor would increase the gap. In the other case, the same variable would reduce the gap. To overcome that, model (2) will be estimated (column 2) using the absolute value of the gap as dependent variable as follows.

$$|Y_i| = \beta \mathrm{X_i} + \varepsilon_i \qquad (2)$$

This specification allows to interpret the results in terms of closing or increasing the gap depending on the sign of coefficient. Nevertheless, the regression (2) is not able to distinguish the factors that influence exclusively underestimation or overestimation cases. To take that into account, we complete this analysis by a multinomial logistic regression (model 3) to identify critical variables explaining the underestimation or overestimation situations relatively to good estimation situation (a relative gap between -10% and 10%). The results of this regression are shown in the columns (3) and (4) of regression results tables.

$$\ln \frac{\Pr(Z=m)}{\Pr(Z=1)} = \beta_m \mathrm{X_i} \quad \text{with m=2 or 3} \qquad (3)$$

where X is the same as in model (1) and the dependent variable Z has three levels such as:

$$Z = \begin{cases} 1\text{: Good estimation: if } -10\% \leq RB_i \leq 10\% \\ 2\text{: Underestimation: if } RB_i > 10\% \\ 3\text{: Overestimation: if } RB_i < -10\% \end{cases}$$

where RB_i is the relative bias.

3.2. Empirical results

Tables 17 to 20 report results from the three models estimated for each of the four country in the sample (Burkina Faso, Mali, Niger, and Nigeria). The dependent variable (DV) of the first model (estimated using OLS) is the observed bias (GPS - SR); in the second one, estimated using OLS too, the DV is the absolute value of the bias; and the third model is a Multinomial Logit using the three-categorical variable associated to the quality of self-reporting (the reference being good estimation).

[INSERT TABLE 17]

[INSERT TABLE 18]

[INSERT TABLE 19]

[INSERT TABLE 20]

In the first specification, it appears that not all variables are influential in all context. Main factors include respondents' age, education, gender, ethnic group, property right, plot relief, extension services, and plot size. In Nigeria, the area unit used (heaps, stands, ridges, plots, acres, hectare, or squared meter) by farmers for plots area size estimation is found also to be essential in explaining the bias. Respondent's age has a U-shape relationship with the bias in Burkina Faso, Niger, and Nigeria, indicating that the bias is greater for older respondents. This result is in line with that of Carletto et al. (2013) in Uganda context. That variable has no impact in Mali context. The plot size has a strong and nonlinear (U-shape) relationship with the observed bias in Burkina Faso, Niger, and Nigeria, but a positive and linear effect on the bias in Mali.

When the absolute value of the bias is considered, in the second specification, the relationship with the respondent's age is significant and quasi-linear (the squared term's coefficient is too small) only in Nigeria context. The plot size is significant for all countries. Larger bias is reported for large-scale plots. Results show also that education is crucial to closing the reporting error in Niger. In Nigeria, respondents that use an area unit other than standard units (acres, hectare, and squared meter) report better than others. In Mali, there is a significant relationship between respondent's gender and the reported bias. Female reporters tend to have higher bias than male reporters. In addition, the land's acquisition status (Mali) and plot's relief (in Mali and Burkina Faso) are also relevant for the quality of the plot size estimation.

Looking at the third specification, results show that factors affecting large underestimation or overestimation risk vary enough across countries. In Burkina Faso, the only factor that really discriminates large positive bias (underestimation cases) is the plot size. The relative risk ratio (RRR) associated to this variable shows that the risk of underestimation has an inverted U-shape, which means that the risk of underestimation is lower for large plots. On the other hand, the risk of overestimation is significantly influenced by the respondent's age, the plot's relief (topography), and the plot size. The expected risk of overestimation in Burkina Faso is higher when the plot's relief is not a plain. The plot size and the respondent's age also have a strong and nonlinear relationship (U-shape) with the risk of overestimation. Younger respondents have a lower risk of over-reporting while older respondents tend to increase the risk overestimation. Similarly, the risk of overestimation is lower for small-scale plots and higher for large-scale plots.

In Mali, the respondents' age is quasi-linear and decreases the risk of underestimation by about 3%, while it has no effect on the risk of overestimation. On the other side, the plot size increases the risk of underestimation by 13.6% but decreases the risk of over-reporting by 40%. In contrast to all expectation, results show that the plot's relief, education, and gender have no impact on overestimation nor on underestimation. Other factors affecting underestimation or overestimation risk include land's acquisition status, locations, and ethnic group.

In Niger, the risk of underestimation is affected by gender, region, and plot size. Female reporters increase the risk relatively to male reporters. In addition, larger plot size increases the risk of underestimation by 21.3%. On the other hand, the risk of overestimation is significantly influenced by the respondent's age, locations, and plot size. Each additional year of the respondent increases the risk of overestimation by 4%, whereas one additional hectare of land decreases this risk by about 32.2%.

In Nigeria, unit of the measurement used by reporters, extension services, education and plot size are the main factors that discriminate large bias. Attending primary school relatively to no education decreases the risk of overestimation by around 25%. The access to the extension services decreases the risk of underestimation by 59%. Plot size has a significant nonlinear relationship with the quality of reporting the area size. The risk of underestimation is increased for small-scale plots but it is decreased for larger plots. The opposite effect is observed for the risk of overestimation.

4. Conclusion and recommendations

The objectives of this paper are to assess the importance of plot land area estimation bias by farmers in West African countries context and to identify factors that influence acreage discrepancy between farmers' self-reported and GPS measured plot size. The study focuses on four countries, namely Burkina Faso, Mali, Niger, and Nigeria. Planted area estimation of a country is essential for agricultural policies. Therefore, it is crucial to have a clear idea of how well farmers estimate their planted areas. Data used in this paper are the recently collected LSMS-ISA data which includes GPS and self-reported land areas for the considered West African countries.

The analysis reveals a systematic discrepancy between GPS and self-reported land areas for all four countries. On average, reporters overestimate land areas by 14% in Burkina Faso, 18% in Niger, and 171% in Nigeria. Unlike these countries, in Mali, respondents tend to under-report the plot size by 40% on average. Therefore, the bias of land areas estimation is important in all those countries, but the extent is very diverse from one country to another. Results also show a consistent impact of the plot size on the quality of the estimation. As found by De Groote and Traoré, (2005), Carletto et al. (2013, 2015a), small plot areas are generally over-reported by farmers while larger plot areas are systematically under-reported. For example, in Burkina Faso, small-scale plots (less than 1 ha) which account for 69% of the plots are overestimated by around 77%, while larger plots (more than 3 ha, 5% of the plots) are underestimated by 31%. In Mali, smaller plots (56%) are over reported by 75% and larger plots (17%) are underestimated by 75%. In Niger, the results are similar, small plots reporters (40%) overestimate by 153% and large plots holders (21%) underestimate by 28%. In Niger, the observation is the same, small-scale plots (89%) are largely over reported by 312%, while large-scale plots are under-reported by 40%. On the other hand, the proportion of good estimation varies from 9% in Nigeria to 28% in Mali. Therefore, it is crucial for policy-makers to think about this issue more seriously.

The source of misreporting varies immensely from country to country. For all countries, larger farms tend to make larger errors on their land size than those who have small farm sizes. In Nigeria, the quality of farmers' area estimation depends also on the area unit used (standard or nonstandard). In Mali, the mode of land acquisition plays a key determinant role of the reporting accuracy. As Carletto et al. (2015) plot's relief is relevant in Mali and Burkina Faso to closing area gap.

Our results show that the risk of large underestimation or overestimation is not affected by the same factors. In other words, underestimation and overestimation are not symmetrical events. For all countries, an evidence of a nonlinear relationship between the plot size and the risk of overestimation (underestimation) are found. Indeed, small-scale plot increases the risk of underestimation, while larger plot decreases the risk and the opposite effect is observed for the risk of overestimation. Respondent's age (in Burkina and Mali), land acquisition status (in Mali), gender (in Niger), area unit measure and extension service (in Nigeria) all matter in explaining the underestimation risk. On the other hand, respondent's age affects the risk of overestimation in Burkina Faso, Niger and Mali. Unlike other countries, in Nigeria, the education and the area unit of measurement are the factors that control the risk of overestimation.

Bibliography

Carletto, C., Gourlay, S., Murray, S., & Zezza, A. (2015, March). Welcome to Fantasyland: Comparing Approaches To Land Area Measurement In Household Surveys. In *2015 Conference, August 9-14, 2015, Milan, Italy* (No. 211849). International Association of Agricultural Economists.

Carletto, C., Gourlay, S., & Winters, P. (2015). From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, *24*(5), 593-628.

Carletto, C., Savastano, S., & Zezza, A. (2013). Fact or artifact: the impact of measurement errors on the farm size–productivity relationship. *Journal of Development Economics*, *103*, 254-261.

De Groote, H., & Traoré, O. (2005). The cost of accuracy in crop area estimation. *Agricultural Systems*, *84*(1), 21-38.

Holden, S. T., & Fisher, M. (2013). Can area measurement error explain the inverse farm size productivity relationship?. *Norwegian University of Life Sciences, Ås, Norway*.

Keita, N., & Carfagna, E. (2009, August). Use of modern geo-positioning devices in agricultural censuses and surveys: Use of GPS for crop area measurement. In *Bulletin of the International Statistical Institute, the 57th Session, 2009, Proceedings, Special Topics Contributed Paper Meetings (STCPM22), Durban*.

Kilic, T., Carletto, C., Zezza, A., & Savastano, S. (2013). Missing (ness) in action: selectivity bias in GPS-based land area measurements. *World Bank Policy Research Working Paper*, (6490).

Muwanga-Zake E.S.K. and J.B. Magezi-Apuuli. (2005). Experience with GPS Equipment in Measuring Crop areas: The Case of Uganda. *African Statistical Journal.* Vol1.

Palmegiani, G. (2009). Measuring cultivation parcels with GPS: a statistical evidence. *The Wye Group Handbook: Statistics on Rural Development and Agriculture Household Income*.

Schoning, P., Apuuli, J. B. M., Menyha, E., Zake-Muwanga, E. S. K. (2005). “Handheld GPS equipment for agricultural statistics surveys: Experiments on area-measurement and geo-referencing of holdings done during fieldwork for the Uganda Pilot Census of Agriculture.” Statistics Norway Report 2005/29.

Annex

Table 17: Determinants of land measurement bias in Burkina Faso

Dependent variable	OLS		Multinomial Logit	
	Bias [GPS - SR] (ha)	|Bias|	RRR	
			Bias > 10%	Bias < -10%
Gender	0.081**	-0.037	1.201	1.104
Age	0.012*	-0.005	0.953*	0.932***
Age squared	-0.000*	0.000	1.001*	1.001***
Education				
Primary	0.078	-0.052	0.978	0.697
Secondary and plus	0.412***	0.060	1.460	1.033
Property	0.045	-0.023	1.127	1.150
Plot relief	-0.073*	0.065*	0.960	1.388*
Extension service	0.058	-0.017	1.148	1.084
Region				
Boucle du Mouhoun	-0.203**	0.003	0.544*	0.411***
Sud Ouest	0.197***	-0.200***	0.560**	0.265***
Centre Ouest	0.275***	-0.218***	0.597*	0.253***
Plateau Central	0.059	-0.073*	0.528**	0.347***
Hauts Bassins	-0.192	0.353	0.939	0.609
Centre	0.289***	-0.242***	0.812	0.263***
Cascarde	-0.266**	0.062	0.414***	0.558*
Centre Sud	0.162**	-0.155**	0.843	0.319***
Plot size (GPS)	0.213***	0.222***	1.177**	0.300***
Plot size (GPS) squared	0.012***	0.012***	0.998**	1.019***
Constant	-0.853***	0.476***	5.834**	96.594***
Observations	2,962	2,962	2,962	2,962
Adjusted (Pseudo) R-squared	0.706	0.786	0.129	0.129

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 18: Determinants of land measurement bias in Mali

	OLS		Multinomial Logit	
Dependent variable	Bias [GPS - SR] (ha)	|Bias|	RRR	
			Bias > 10%	Bias < -10%
Gender	0.296***	0.212**	1.015	1.039
Age	-0.003	0.010	0.964***	0.995
Age squared	-0.000	-0.000	1.000**	1.000
Education				
Primary	0.156	-0.181	1.051	1.016
Secondary and plus	0.208	-0.071	1.213	0.853
Property	0.249***	0.218***	1.180*	0.898
Plot relief	0.492**	-0.445**	1.132	1.061
Ethnic				
Peulh/Foulfoulbe	0.234**	0.333***	1.143	1.200
Sarakolé	-0.292	0.116	1.438***	1.464***
Sénoufo/Minianka	-0.609***	0.070	0.894	1.010
Dogon	0.520***	0.680***	2.139***	1.650***
Bobo/Dafing/Samogo	0.118	0.225	2.498***	1.806***
Autres ethnies du Mali	-0.106	0.474*	1.357**	1.597***
Region				
Koulikoro	0.299**	0.229*	1.895***	2.576***
Kayes	0.526***	0.391**	1.243*	1.599***
Ségou	-0.103	-0.032	1.500***	1.681***
Mopti	-0.437**	-0.199	1.130	3.341***
Tombooutchou	0.627*	-0.078	1.474	1.094
Gao	0.435	0.336	0.692	1.565
Bamako	0.013	-0.197	0.617	0.000***
Plot size (GPS)	0.826***	0.809***	1.136***	0.599***
Plot size (GPS) squared	0.001	0.001	0.999***	1.005***
Constant	-1.559***	-0.492	1.529	0.942
Observations	8,223	8,223	8,223	8,223
Adjusted (Pseudo) R-squared	0.819	0.823	0.0961	0.0961

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 19: Determinants of land measurement bias in Niger

Dependent variable	OLS		Multinomial Logit	
	Bias [GPS - SR] (ha)	|Bias|	RRR	
			Bias > 10%	Bias < -10%
Gender	0.052	0.141	1.445**	1.252
Age	-0.046**	-0.008	1.011	1.044*
Age squared	0.001***	0.000	1.000	1.000*
Head's education				
Primary	0.580***	-0.388**	1.210	0.674
Secondary and plus	0.704***	-0.631***	0.630	0.682
Ethnie				
Haoussa	0.435	-0.051		
Kanouri-Manga	0.367	0.007		
Touareg	0.154	-0.081		
Other	0.442	-0.069		
Property right	-0.070	0.097	0.940	1.230
Plot relief	-0.117	-0.021	1.048	1.043
Region				
Diffa	0.681***	-0.532***	2.104***	0.568**
Dosso	0.311*	-0.175	1.217	1.048
Maradi	0.337**	-0.220*	1.076	0.989
Tahoua	0.117	0.130	0.922	1.017
Tillabéri	-0.501	0.753**	0.990	0.764
Zinder	1.566***	-0.542*	1.097	0.279***
Niamey	0.036	-0.877	0.594	0.369
Plot size (GPS)	0.530***	0.456***	1.213***	0.578***
Plot size (GPS) squared	0.004***	0.005***	0.998***	1.005***
Constant	-1.136**	0.897*	0.991	3.886**
Observations	3,417	3,417	3,417	3,417
Adjusted (Pseudo) R-squared	0.350	0.372	0.131	0.131

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 20: Determinants of land measurement bias in Nigeria

	OLS		Multinomial Logit	
Dependent variable	Bias [GPS - SR] (ha)	|Bias|	RRR	
			Bias > 10%	Bias < -10%
Gender	-0.159	0.139	1.280	1.022
Age	-0.057**	0.052**	0.979	0.988
Age squared	0.000**	-0.000**	1.000	1.000
Education				
Primary	0.062	-0.057	0.997	0.744**
Secondary and plus	-0.103	0.073	0.751	0.903
Poperty	-0.306	0.376	1.179	0.864
Slope	-0.083	0.084	0.968	1.001
Extension	-0.238	0.100	0.411***	1.065
Unit of measure				
Heaps	2.734***	-2.308***	7.939***	0.772
Ridges	0.509*	-0.371	6.212***	1.549**
Stands	2.790***	-2.267***	11.085***	0.651**
Plots	1.601**	-0.990	16.213***	0.307***
Zone				
NORTH EAST 1,213	0.994***	-0.937***	0.365***	0.489***
NORTH WEST	1.319***	-1.184***	1.107	2.088***
SOUTH EAST	-0.245	0.135	0.777	0.481***
SOUTH SOUTH	-0.806	0.737	0.962	1.011
SOUTH WEST	0.846***	-0.925***	0.638	0.561**
dist_popcenter2	-0.011	0.008	0.991	0.993
Plot size (GPS)	0.515***	0.059	2.890***	0.163***
Plot size (GPS) squared	0.037**	0.053***	0.928**	1.165***
Constant	-0.698	0.482	1.026	45.585***
Observations	5,076	5,076	5,076	5,076
Adjusted (Pseudo) R-squared	0.0412	0.0315	0.220	0.220

Note: *** p<0.01, ** p<0.05, * p<0.1