Accuracy of agricultural data and implications for policy : Evidence from maize farmer recall surveys and crop cuts in the Guinea Savannah zone of Ghana

Agricultural Systems

Adzawla, William; Setsoafia, Edinam D.; Setsoafia, Eugene D.; Amoabeng-Nimako, Solomon; Atakora, Williams K. et al

https://doi.org/10.1016/j.agsy.2023.103817

Agricultural Systems 214 (2024) 103817

Contents lists available at ScienceDirect

Agricultural Systems

journal homepage: www.elsevier.com/locate/agsy

Accuracy of agricultural data and implications for policy: Evidence from maize farmer recall surveys and crop cuts in the Guinea Savannah zone of Ghana

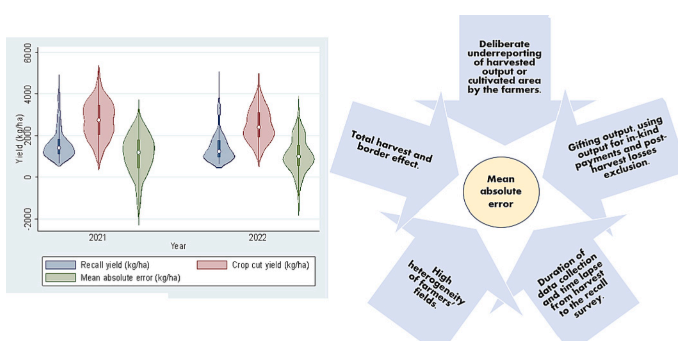
William Adzawla[a,*], Edinam D. Setsoafia[b], Eugene D. Setsoafia[c,d], Solomon Amoabeng-Nimako[c,d], Williams K. Atakora[a], Prem D. Bindraban[a]

[a] International Fertilizer Development Center, Accra, Ghana
[b] University for Development Studies, Department of Agribusiness, Nyankpala, Tamale, Ghana
[c] Wageningen University & Research, Centre for Crop Systems Analysis, P.O. Box 430, 6700 AK Wageningen, the Netherlands
[d] Agricultural Innovation and Technology Transfer Center, Mohammed VI Polytechnic University, Ben Guerir 43150, Morocco

HIGHLIGHTS

- Precise data is a prerequisite for policy design towards agricultural development and transformation.
- Maize yield measured through crop cut is significantly higher than those measured through farmer recall surveys.
- The interpretation of farm yield data should be properly situated within its given context.
- Yield measurement error are explained by cultivated area and socioeconomic conditions of the farmers.
- Crop cut yield estimation is costly but has limited biases.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Dr. Jagadish Timsina

Keywords:
Agricultural data
Crop cut
Data accuracy
Farmer recall
Maize yield
Measurement error

ABSTRACT

CONTEXT: The need for accurate data in policy design aimed at agricultural transformation cannot be overemphasized. Unfortunately, the relevance of agricultural research in addressing the needs of farmers has been questioned due to debates about appropriate methodologies and approaches for establishing research activities and, in other instances, poorly reasoned premises and paltry delineation, definition, and understanding of the system being studied. For a country like Ghana, where agricultural transformation is a prerequisite for its sustainable development, an understanding of the accuracy of farm data measurement is necessary.
OBJECTIVE: The objective of this study is to estimate the yield measurement error and to analyze the sources of such measurement errors among the farmers of the Guinea Savannah zone of Ghana.
METHODS: Two years' data for both farmer recall surveys and crop cuts were used. Descriptive statistics, regression and sensitivity analyses were done to achieve the objectives of the study.
RESULTS AND CONCLUSIONS: On average, farmers' recall of maize yields (1544.6 kg/ha) was lower than the estimated crop cut yields (2593.9 kg/ha), although about 11.2% of the farmers recalled higher yields than their

* Corresponding author.
E-mail address: wadzawla@ifdc.org (W. Adzawla).

https://doi.org/10.1016/j.agsy.2023.103817
Received 6 July 2023; Received in revised form 25 November 2023; Accepted 26 November 2023
Available online 5 December 2023
0308-521X/

estimated crop cut yields. The estimated average percentage error in yield measurement between crop cuts and farmer surveys was 36.4%. These yield measurement errors are due to systematic biases, including those involving the recall of farm size, and the socioeconomic conditions of the farmers. Although a crop cut is costly, it has limited bias in providing a better measure of yield than farmer recall surveys. Irrespective of the method used, however, more attention should be given to potential sources of systematic bias in the design and data collection. Moreover, for proper interpretation, yield estimates from recall surveys and crop cuts should be properly interpreted as economic yield and biological yields, respectively.

SIGNIFICANCE: This paper provided clarity on the differences in maize yield estimates in Ghana and provided measures on how to improve precision in yield data.

1. Introduction

The development of agriculture plays a vital role in enhancing food security, alleviating poverty and inequality to foster shared prosperity, and establishing supply-demand chain systems, such as the provision of raw materials to industries. The agriculture sector of Ghana contributed about 21% to its Gross Domestic Product (GDP) in 2022 (GSS, 2023) and formed a core component of rural livelihoods (Adzawla et al., 2022; Giller, 2020). Unfortunately, Ghana has not been able to harness its full agricultural potential. Yields of major crops remain below the potential yields, and farmers continue to be the poorest people in Ghanaian society. Successive governments have rolled out various policies, programs, and interventions with the goal of modernizing agriculture, achieving greater impacts, and improving the livelihoods of smallholder farmers. An important requirement in the design of such agricultural policies supporting agricultural development is precise agricultural data (Ambler et al., 2021; Kosmowski et al., 2021; Wineman et al., 2019; Blandford, 2007). Thus, a precise understanding of agricultural livelihood activities of farm households is vital for designing an effective policy (Ambler et al., 2021), assessing the suitability of low-emission agricultural activities under various production environments (Sapkota et al., 2016), evaluating the performance of new technologies (Kosmowski et al., 2021), and monitoring progress on achieving development goals, such as Sustainable Development Goal 2 (SDG 2) – zero hunger (Lobell et al., 2019).

There is a general understanding of the relevance of smallholder agriculture in improving economic conditions, but there is near negligence in ensuring quality statistics for monitoring the progress of agricultural development initiatives (Carletto et al., 2015). Accordingly, obtaining precise and timely estimates on production inputs, production volumes, and yields remains a significant challenge, especially in smallholder agricultural farming systems. Although the shortfall in the accuracy of agricultural data is not new, nor isolated to only sub-Saharan Africa (SSA) (Wahab, 2020; Carletto et al., 2015), it is extremely worrisome for a country such as Ghana, where agricultural development and transformation form a major determinant for the socioeconomic development of the country. This lack of reliable and accurate data limits a proper understanding of the actual performance of the agricultural sector and affects the design of appropriate policies and interventions to improve the living conditions of farmers. In addition, the quality and relevance of research on agricultural systems is low due to the absence of relevant data (Wineman et al., 2019). For instance, an editorial release (Nature Plants, 2020; Nature, 2020) revealed that not only do most published articles use unoriginal data, but >95% of the 100,000-plus reviewed articles do not provide relevant information on the needs of smallholder farmers (i.e., food security and poverty eradication).

For estimating crop yields, several methodologies have been developed and used, each with their own merits and limitations. The choice of methodology is influenced by several factors, including the required level of precision, resource availability, and the scale of estimation (Sapkota et al., 2016). But generally, agricultural data has been collected by asking farmers and other value chain actors or their agents to recall information on their production activities in previous seasons, including the type and quantity of investments (e.g., cultivated land area, inputs used) and the associated activity outcomes (e.g., total harvest or farm income). This method is particularly common in SSA countries due to its cost-effectiveness, convenience, and suitability for the diverse farming systems (Wahab, 2020). However, there are biases from farmer surveys (Beegle et al., 2012). Although rural farm data has historically been perceived as poor quality, especially if it has not been obtained from a specialized agricultural survey (Beegle et al., 2012), the resultant measurement error has been found to differ based on the estimation method. These potential biases have resulted in a shift in the method for estimation of agricultural statistics, such as crop yields, from farmer recall surveys to other methods, such as crop cuts, as recommended by the Food and Agriculture Organization of the United Nations (Kosmowski et al., 2021; Desiere and Jolliffe, 2018). Although the crop cut methodology has increasingly become a gold standard of measuring crop yield, it does have its limitations; the method only estimates the biological yield and not the economic yield (Fermont and Benson, 2011), the plot cuts are random, and the heterogeneity of farms makes the crop cut yields sensitive to the size of the plot cuts (Sapkota et al., 2016; Fermont and Benson, 2011). Moreover, crop cut estimates also suffer from systematic biases.

Past research has relied on self-reporting methodology to estimate crop yields in Ghana. This has been used widely in gathering socioeconomic data by developmental programs, including the Fertilizer Research and Responsible Implementation (FERARI) program (https://ifdc.org/projects/fertilizer-research-and-responsible-implementation-ferari/). In addition to surveys, the FERARI program has also conducted crop cuts. Yield estimates from the surveys and crop cut data have varied, raising concerns about which method provides the correct information on the observed maize yields on farmers' fields within the program's implementation area. Therefore, it has become important to analyze in detail the yield variations and identify reasons for such possible measurement errors in the yield estimates within the Ghanaian context. This study is also important because of a lack of empirical evidence in comparing yield estimates under different methodologies in Ghana and many other parts of sub-Saharan Africa. Finally, this study examines the appropriateness and relevance of the conclusions and recommendations from previous studies on the basis of the quality of the data used.

2. Methodology

2.1. Study location

The data was gathered in the Guinea Savannah agroecological zone of Ghana, specifically five northern regions – Northern, North East, Savannah, Upper East, and Upper West (Fig. 1). The Guinea Savannah zone represents about 62% of Ghana's total land area (MoFA, 2021). The soils of northern Ghana are mostly shallow and prone to erosion; they have low to moderate nutrient levels, moisture, and nutrient-holding capacities. Nonetheless, the soils are generally suitable for cereal crop production (Tetteh et al., 2016). Around 78–82% of households in the region are involved in agriculture (GSS, 2019; MoFA, 2021), and about 39.1% of Ghana's starchy staple producers are in the northern savannah

(GSS, 2019a). Farmers in these regions largely depend on their farm returns for providing their food and non-food needs. As a result of poor households' high engagement in agriculture, coupled with the implementation of several agricultural projects and programs, agricultural research in the regions has gained wide attention.

Maize is one of the major crops cultivated in the Guinea Savannah zone. Maize is not only a major food crop for the farmers, but also a cash crop for meeting the income needs of the farm families. Except for the Upper West Region in 2019 and 2020, average yields for the northern regions have been below the national maize yield average in the last few years (Fig. 2). Overall, huge potential exists for increasing yields in the northern regions, as the average yield fell below the achievable yield of 5.5 mt/ha (MoFA, 2021), while crop modeling reveals the yield potential here to be 6–8 mt/ha (Boullouz et al., 2022). The major factors affecting maize yield in this area include poor soil conditions and climate variability (Kouame et al., 2023).

2.2. Data collection

Two forms of primary data were collected for both 2021 and 2022: farmer recall surveys and crop cuts. In each year, the crop cuts were done first, and the same farmers were subsequently interviewed to recall their production figs.

2.2.1. Farmer recall surveys and yield estimation procedure

In 2021, a multistage procedure was used for sampling the farmers. In the first stage, purposive sampling was used to select farmers who used any form of fertilizer during the 2021 cropping season. This was necessitated by the introduction of ammonium sulfate into the government of Ghana's Planting for Food and Jobs program (https://mofa.gov.gh/site/programmes/pfj) to complement the shortfall in urea imports into the country. Therefore, an initial list of 280 farmers was generated to determine the various fertilizer types that they used for topdressing. At the second stage of the sampling procedure, the farmers were placed in four categories: those who used urea, NPK compound or blended fertilizers, ammonium sulfate for topdressing, and those who did no topdressing. In the third stage, an unequal sample was drawn from each of the four categories of farmers using a simple random sampling procedure. Overall, 181 farmers were successfully interviewed in 2021, about nine months after harvesting their maize.

In 2022, a multistage sampling procedure was used for selecting the farmers. In the first stage, stratified sampling was used to put the districts in each region into different classes based on their historical yield levels. The groups included low-yielding (≤2 mt/ha), medium-yielding (2–3 mt/ha), and high-yielding (>3 mt/ha) districts. In the third stage, one district from each stratum was selected from each of the five regions using a simple random sampling procedure. With the help of agricultural extension officers, simple random sampling was used to select four maize-producing communities in each selected district. Finally, simple random sampling was used to select 311 farmers who were interviewed about two months after harvest.

The recall data were collected through face-to-face interviews using a questionnaire. The questionnaire was transcribed onto a mobile platform and pretested for reliability and validity, and to test the practicability of the mobile data collection procedure. The data collection tools were finalized based on the pretest results. Research assistants were

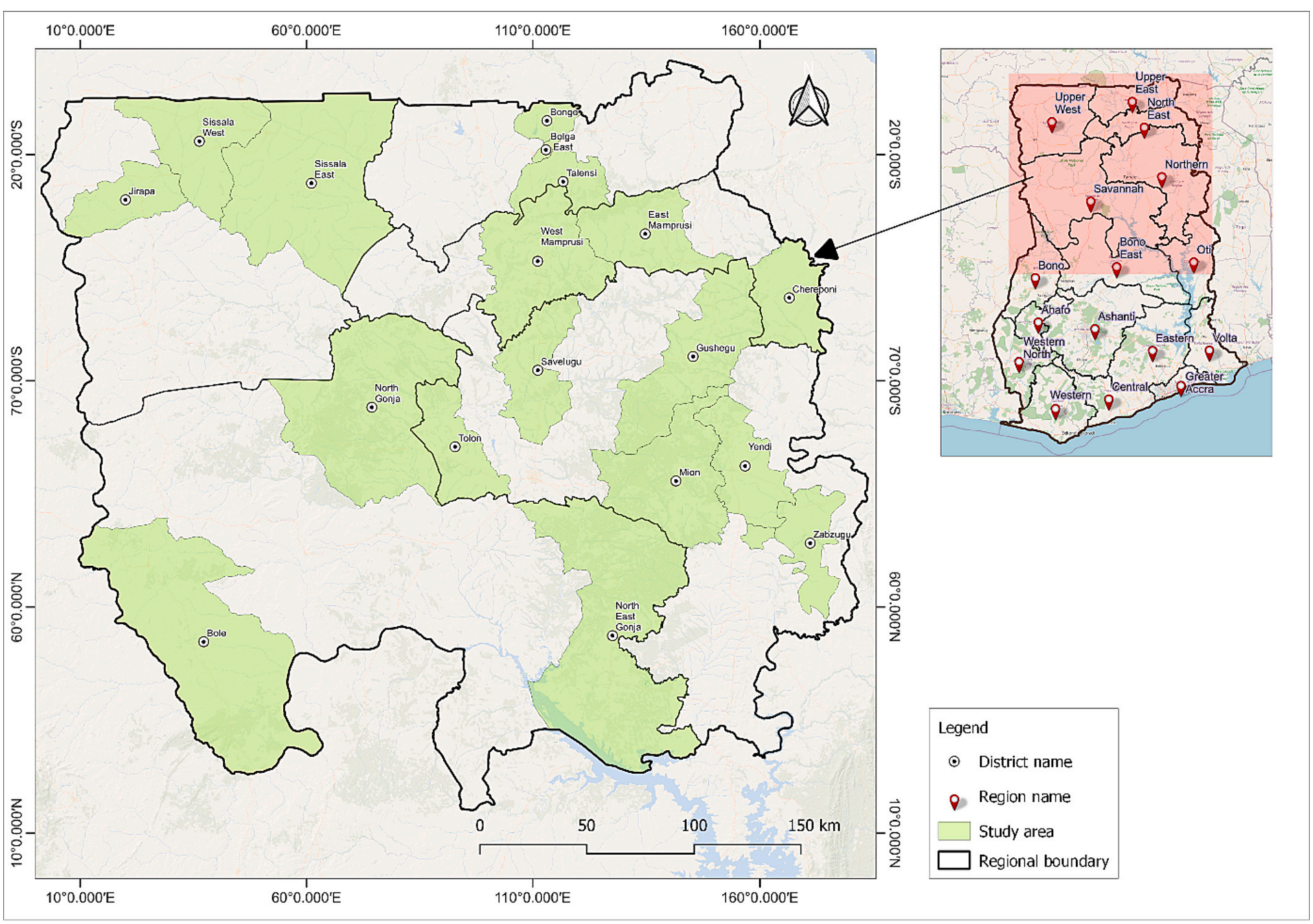


Fig. 1. Study area.

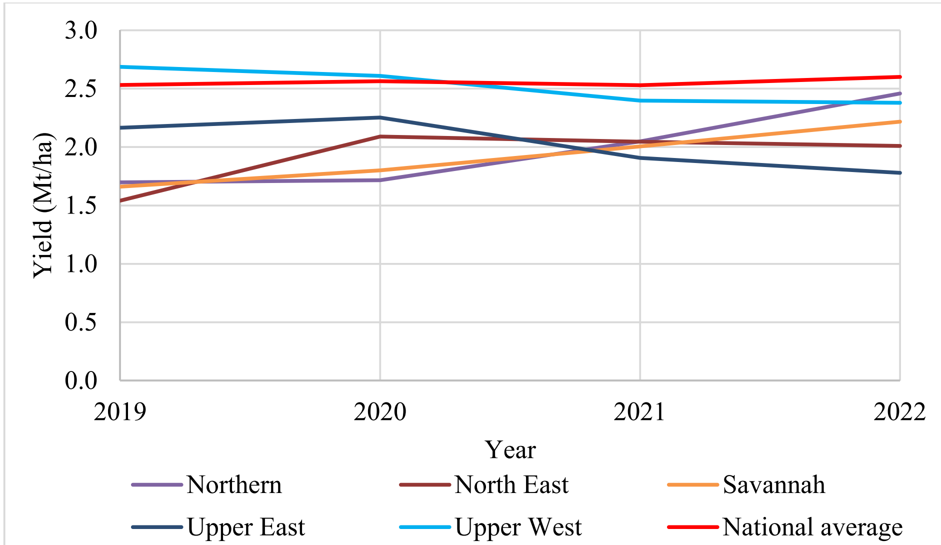


Fig. 2. Maize production in the study regions.
Source: Ministry of Food and Agriculture (MoFA) Statistics, Research and Information Directorate (SRID) (unpublished).

recruited based on their experience in survey data collection and were assigned to different regions based on their understanding of the farmers' local language. The research assistants were trained on the objectives of the survey and use of the mobile data collection app. The data included the socioeconomic characteristics of the farmers; production investments; quantity of maize output realized; and the total cultivated area. The summary of how farm size (step 1) and farmer recall yields (steps 2–4) were measured are as follows:

1. The farm size used in this study was measured through farmers' recalls. The farmers were asked 'how many acres did you cultivate to maize in the last production season?' The reported farm size was then converted from acres to hectares by dividing the individually reported figures with a conversion factor of 2.471.
2. The farmers were asked 'how many bags of maize did you obtain from the total cultivated area and what is the unit size (50kg or 100kg) of the bag?'. This resulted in the output of the farmer.
3. The output from step 2 was converted to its kilogram equivalent by multiplying the number of bags with 50 or 100 depending on the reported unit size of bag.
4. The farmer's yield was then estimated by dividing the total kilograms of harvested output by the total number of hectares from which the output was realized.

2.2.2. Maize crop cuts

Crop cuts were conducted on the same farmers' fields by field research officers with much experience in yield estimation. A protocol detailing the parameters to perform the measurement was designed for crop cut data collection. Two research officers, together with the maize farmer and two laborers, were involved in the data collection on each farm. The crop cut yield data were estimated through the following steps:

1. The harvest from each farm was conducted on three sections, measuring 2 × 2 m (4 m^2) each. The sections were selected by physically observing the heterogeneity in the maize crops on the farm. The three sections included (1) one area with potential high yield, (2) one area with potential low yield, and (3) one area with a potential yield between the first two sections.
2. The weight of the harvested output (cobs with grains) of each harvested section of the field was taken and multiplied by 10,000 (note: 10,000 m^2 = 1 ha). Since each data spot measured 4 m^2, the result was divided by four to obtain the estimate in kilograms per hectare.
3. A shelling percentage of 80% was applied to the weight of the cob with grains to arrive at the weight of the grains. Thus, 0.8 multiplied by the result in step 2. The 80% shelling percentage applied in this study is due to its wide recommendation for normal maize ears at harvest (Lauer, 2002; Tandzi and Mutengwa, 2020).
4. The percentage moisture content of the grains was taken for each sample using a moisture meter and the results were divided by 100 to obtain the proportion of grain to moisture. Averagely, the moisture content of the grains was 14.1%. The results were multiplied by the fresh grain weight at step 3 to obtain the moisture content of the grains in kg.
5. The dry grain weight for each sample was then obtained by subtracting the moisture content (step 4) from the fresh grain weight (step3).
6. Finally, the dry grain yield for each farm was obtained by calculating the average result from the three sections.

2.3. Data analysis

The data were analyzed using descriptive statistics, as well as multiple linear and logit regressions. Descriptive statistics were used to estimate the means and to explore the distribution of the yield and measurement errors using histogram, scatter, and violin plots. The strength of the relationship between the yield estimation methods was assessed using Pearson's correlation coefficient. The systematic yield measurement error was estimated using the mean absolute error (MAE) and the absolute percentage error (APE). The MAE is a measure of error between paired observations (in this case, the crop cut and recall yield estimation methods) determining the same parameter or statistic (in this case, maize yield). This is estimated as:

$$MAE_t = \frac{\sum(y_{ct} - y_{rt})}{n_t} \tag{1}$$

where y_c is the yield obtained by crop cut, y_r is the yield obtained in the farmer recall survey, n is the number of observations, and t is the period under consideration. The APE is estimated as:

$$APE_t = \frac{\sum(y_{ct} - y_{rt})}{y_{ct}} * 100 \tag{2}$$

The higher the MAE or APE, the higher the difference in the yield obtained under the two estimation methods for each farmer. A low MAE or APE implies both yield estimation methods tend to give similar yield values.

Ordinary least square (OLS) and logit regressions were fitted to explain the effect of farm size, other production inputs, and socioeconomic factors on yield measurement error and the effects of production inputs on yield estimates from crop cut and recall surveys. A naïve OLS model, with measurement error using APE as a function of farm size (measured in hectares and classified into four quantiles), was estimated to determine the implications of the cultivated area on the observed measurement error. This is given as:

$$APE = \delta_0 + \delta_1\, farm\ size \quad (3)$$

However, other factors, including the socioeconomic characteristics of the farmers, could explain the variation in the yield measurement error. Therefore, an expanded form of eq. 3 was estimated. Also, to determine the probability that a farmer with a given characteristic will underestimate or overestimate the yield in any given period, a random-effect logit regression was estimated. The basic assumption is that given the binary outcome:

$$y_{it}^{*} = x_{it}^{'}\delta + u_{it} + v_{it};\ i = 1, 2, 3 \ldots n\ and\ t = 1, 2 \quad (4)$$

where y_{it}^{*} is a latent unobserved binary measurement error outcome of a farmer, $x_{it}^{'}$ are the time variant characteristics of the farmers, u_{it} are unobservable time-invariant factors, and v_{it} are unobservable time-variant factors. In the estimation of the random-effect logit model, the u_{it} are assumed to be unrelated to the $x_{it}^{'}$. The observed outcome of y_{it}^{*} is y_{it}, which is defined as:

$$y_{it} = \begin{cases} 1\ if\ y_{it}^{*} > 0 : the\ crop\ cut\ is\ greater\ than\ farmer\ recall\ (underestimating) \\ 0\ if\ y_{it}^{*} \leq 0 : if\ the\ crop\ cut\ is\ lower\ than\ farmer\ recall\ (overestimating) \end{cases}$$

The associated probability of a farmer overestimating the yield is estimated as $P(y_{it} = 1|x_{it}) = F(x_{it}^{'}\delta)$. The empirically estimated logit model is:

$$\begin{aligned} APE(p) = {} & \delta_1 Farm\ size\ quantile + \delta_2 Regional\ location + \delta_3 Farming\ system + \delta_4 Sex + \delta_5 Age + \delta_6 Education + \delta_7 FBO + \delta_8 Experience \\ & + \delta_9 Credit\ access + \delta_{10} Non-farm\ employment + \delta_{12} Fertilizeruse + \delta_{11} Labor + \delta_{11} Year \end{aligned} \quad (5)$$

where δ are marginal effects measuring the change in measurement error (APE) due to change in the factor. $APE(p)$ is defined as 1 if the crop cut yield is higher than the farmer recall yield and 0 if the recall yield is greater than the crop cut yield.

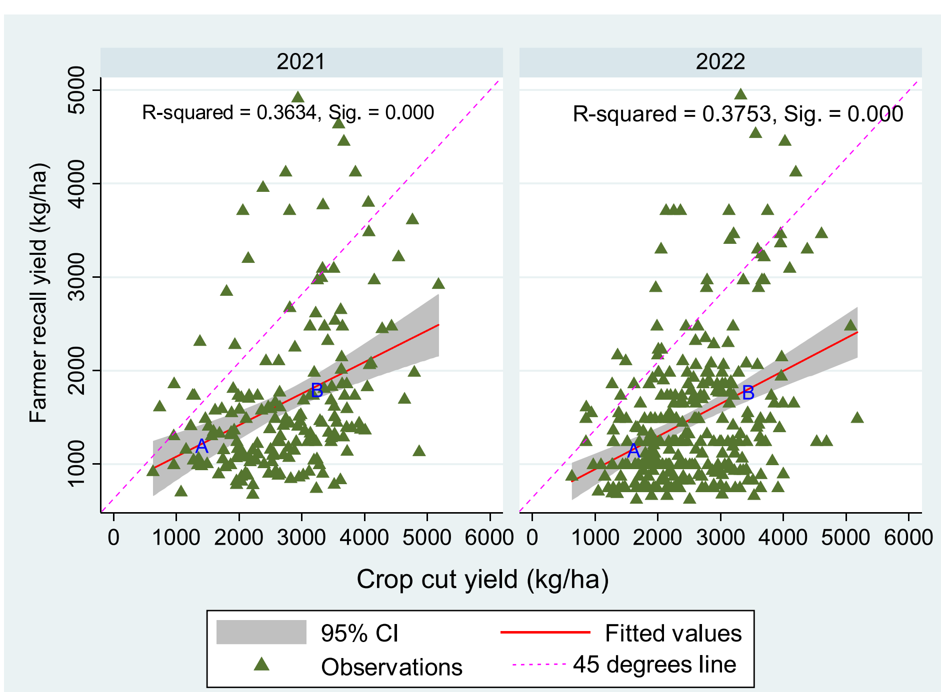


Fig. 3. Scatter plot of yield estimates.

3. Results and discussion

3.1. Descriptive statistics of production and socioeconomic factors

The mean yields from crop cut and farmer recall were 2593.9 kg/ha and 1544.6 kg/ha, respectively, and the average cultivated area was 1.9 ha. The MAE and APE were 1049.3 kg/ha and 36.4%, respectively. The average quantity of local and improved seeds used were 12.3 kg/ha 6.3 kg/ha, respectively. Farmers used about 8 persons/ha for labor and applied about 4 L/ha of herbicide for weed control and about 72% of the farmers applied mineral fertilizers at an average rate of 161 kg/ha. The mean level of formal education was about 5 years, while the mean experience in maize production was about 13 years. About 35% of the farmers belonged to a farmer-based organization (FBO), and 37% of the farmers had access to extension services. Farmers were about 39 years old on average.

3.2. Correlation between farmer recall and crop cut yields

The correlation analysis of the yield estimates (Fig. 3) depicts a similar positive relationship between crop cut and farmer recall yield estimation methods for both years. This indicates that a higher yield estimate under one method is associated with a higher yield estimate under the other. This is confirmed by the significance of the correlation coefficients. For instance, given two farmers, A and B, both the crop cut and recall yield estimates were lower for farmer A than farmer B. However, the fitted lines lie below the 45° (zero measurement error where recall yield equals crop cut yield) line. Thus, the farmer recall yield estimates tend to be lower than the crop cut yield estimates. Generally, 11.2% of the farmers recalled a higher yield than the yield estimate from the crop cut, while 88.8% recalled a lower yield than the crop cut yield estimate. The farmers above the 45° line generally used more fertilizers and improved seeds other than local seeds, relatively younger, mostly members of farmer organizations and accessed extension services. However, the differences in the characteristics between the farmers are statistically insignificant except fertilizer applied and age (Appendix 1). Only about 4% of the farmers had a yield difference that was below a negligible ±100 kg/ha from the 45° line. Paliwal and Jain (2020) also reported that about seven in 10 farmers in Arrah District, India, underreported their yields. Kosmowski et al. (2021) confirmed a positive relationship among yield estimation methods, and the estimated correlation coefficient between farmer estimates and other yield estimation methods was between 0.089 and 0.129. A statistical test of the factors influencing the probability of a farmer underestimating (above the fitted line) or overestimating (below the fitted line) was conducted (Table 3).

3.3. Distribution of estimated yields and measurement error

The violin plots show the box plot and kernel density distribution of each of the yield estimates and MAE (Fig. 4). Generally, the crop cut yield estimates are better distributed with higher average yield than the recall yields. In 2021, the average yields under crop cut and farmer recall were 2743.6 kg/ha and 1667.2 kg/ha, respectively; in 2022, they were 2506.8 kg/ha and 1473.1 kg/ha, respectively. Thus, the MAEs in 2021 and 2022 were 1076.5 kg/ha and 1033.5 kg/ha, respectively. The estimated 2021 crop cut yield compares well with the estimated 2692.3 kg/ha national maize average, as reported by the Food and Agriculture Organization of the United Nations (FAO; https://www.fao.org/faostat/en/#data/QCL), relative to the approximately 2500 kg/ha national average estimated by Statistics, Research and Information Directorate (SRID) of MoFA (MoFA SRID) (unpublished). The average crop cut yields were about 600 kg/ha and 300 kg/ha higher than the regional averages of MoFA SRID in 2021 and 2022, respectively (Fig. 5). Although the MoFA SRID data is also gathered through crop cuts, they average the mean yield for their districts, which could result in the yield gap between the estimated crop cut yield in this study and their yield. In addition, there were significant differences in the yield estimates between the two estimation methods and the FAO and MoFA SRID datasets. Although FAO's official data for 1991–2016 compares well with MoFA SRID national data, FAO's estimated values for 2017–2022 were higher than those of MoFA SRID. The implication is that the yield estimates by various institutions, if not otherwise reported with MoFA SRID data, are not necessarily the same. The regional comparison of yield estimates is shown in Appendix 2.

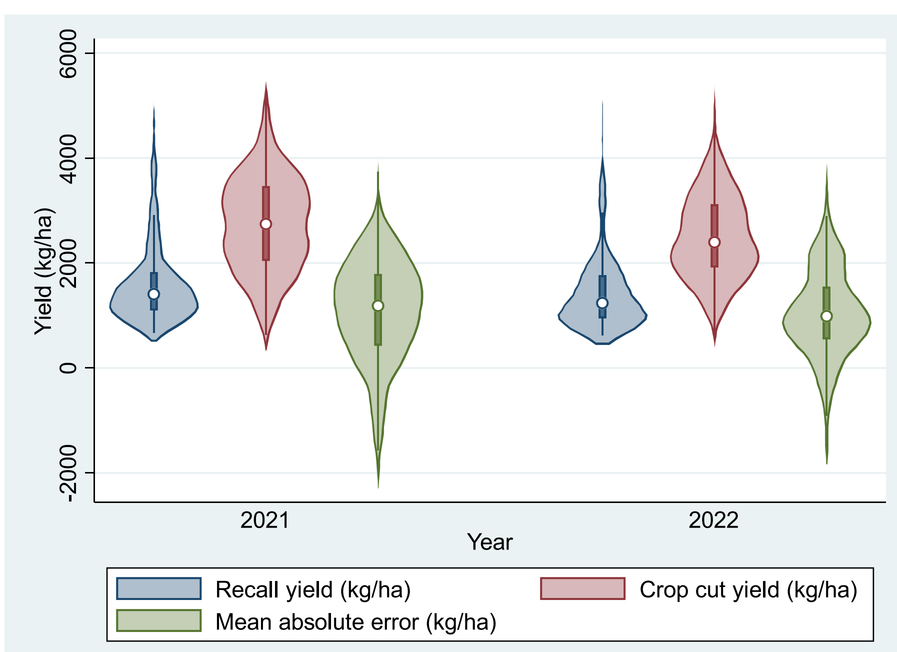


Fig. 4. Plot of mean yield estimates and mean absolute error by year.

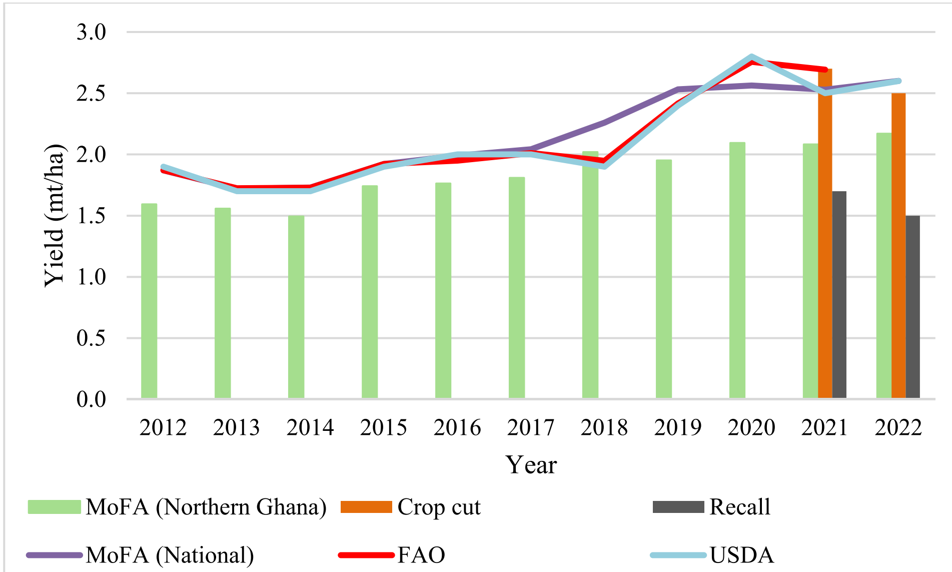


Fig. 5. Yield comparison with other secondary data sources.
Source. Own estimates, MoFA SRID (unpublished), FAOSTAT (https://www.fao.org/faostat/en/#data/QCL) and USDA (https://ipad.fas.usda.gov/countrysummary/default.aspx?id=GH&crop=Corn).

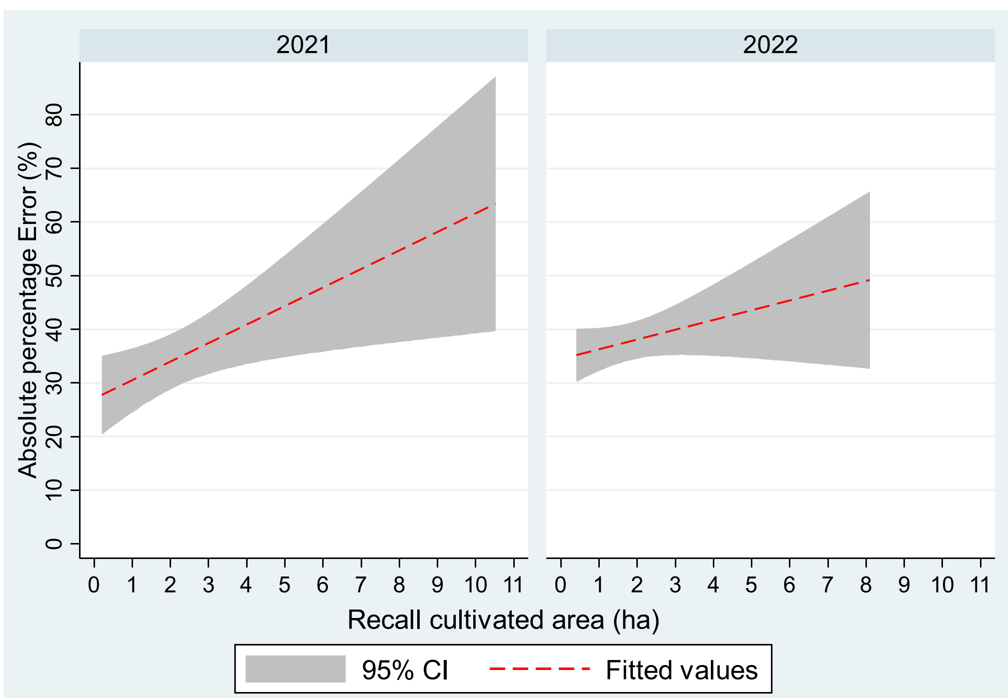


Fig. 6. Relationship between absolute percentage error and cultivated area.

3.4. Absolute percentage error in yield estimates

The estimated APE ranged from −118.7% to 81.3%, with an average of 34.3% in 2021, 37.7% in 2022, and 36.4% overall. The implication is that, although some farmers (11.2%) have overestimated their yields during the recall survey by about 35.4% above the crop cut yield estimates, farmers typically reported lower yields by 45.5% than those estimated through crop cuts. Several studies, including Sapkota et al. (2016), have indicated a high tendency for farmers to overestimate yields under crop cuts due, for instance, to the heterogenous farm conditions and the high tendency to sample plot areas with a good plant stand. While there could be systematic errors in estimating the outputs for both the crop cuts and farmer recall surveys, the latter method also suffered from a systematic error in measuring the farm size, which is the foundation for yield estimation under farmer recall surveys. As seen in Fig. 6, there was a positive relationship between the cultivated area and the APE. Fig. 7 further shows that the mean APE increased from the first quantile to the fourth quantile. The implication is that, as the farm size increases, APE also increases. This is properly discussed in the subsequent sections.

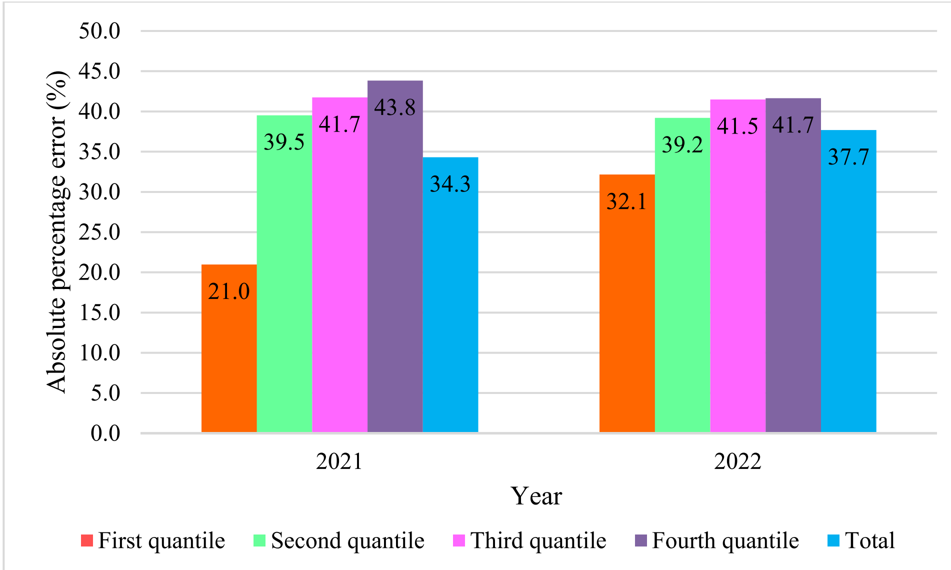


Fig. 7. Absolute percentage error by quantile of cultivated area.
Note: The average farm size for the first, second, third, and fourth quantiles was 0.7 ha, 1.5 ha, 2.4 ha, and 5.5 ha, respectively, in 2021 and 0.7 ha, 1.4 ha, 2.2 ha, and 4.1 ha, respectively, in 2022.

3.5. Regression analysis of the effect of cultivated area and socioeconomic factors on measurement error

Table 2 shows the regression analysis of the relationship between farm size and APE in yield estimation. Both model 1, where cultivated area was measured in hectares, and model 2, where cultivated area was classified into four quantiles, confirm the results presented in Figs. 5 and 6, in which higher measurement errors are due to increasing farm size. Thus, yield cut estimates are very high on larger farms, while such farmers with larger farms tend to underestimate their yield in a recall survey. A third model (Table 3) also shows that, in addition to farm size, other socioeconomic factors explained the direction or magnitude of the measurement error among the farmers. These factors include regional location, age, education, and farming experience. For farm size, farmers within the first quantile had significantly lesser probability of over-reporting their yield than those in other quantiles, consistent with the results of Table 2. Although this is in alignment with other studies (e.g., Fermont and Benson, 2011), Gourlay et al. (2019) estimated that farmers in Uganda who cultivated smaller areas overreported their maize yields. Juxtaposing these with the regional estimates discussed next explains that different groups of people have different perspectives and sentiments in recalling or providing information on their production activities.

The regional location of the farmers influenced the direction of the yield measurement error. Using the Northern Region as a reference group, farmers located in the North East and Upper West regions had significantly lesser probability of underestimating their recall yields, while those located in the Savannah and Upper East regions had insignificantly higher probability of overestimating their recall yields. The mean APE also confirmed that farmers in North East and Upper West regions had an average APE of 32% and 17%, respectively, compared to the Northern Region average of 47%, while Upper East and Savannah regions had an average APE of 51% and 47%, respectively. Historically, maize yields have been higher in the Upper West Region (Fig. 2). Hence, there is less incentive for farmers in this region to underreport their yields; as such, the APE is expected to be less. Previously, Adzawla et al. (2021) recommended that the attributes of farmers in the Sissala East and West districts of the Upper West region should be improved and promoted due to their observed suitable production outcomes. Both age and higher education increase the probability and the magnitude of a farmer underestimating the yield. Although the result for education is surprising, the data show the level of education is not only low but indifferent (4.8 years each) between farmers who underestimated and those who overestimated their yields. Also, the recall ability of younger people is generally high, and they also tend to report higher yields in order not to demean their youthfulness. Gourlay et al. (2019) found that, although insignificant, the directional effect of both education and age on measurement error is based on the classification of the farmers into different percentiles. The results also show that farmers who have cultivated maize for several years are most likely to overestimate their yields. This is particularly due to the influence of the historical information they have on maize yield, which tends to influence their reported production figures. The locational effect on APE suggests the farmers must be carefully quizzed to arrive at the accurate information when conducting recall surveys in those areas. Although insignificant, the time-varying factor, year, which proxied the time lapse between the crop cuts and recalls is negatively related to the measurement error. Thus, the longer the time lapses, the closer the farmers' yield recalls get closer to the yields obtained under crop cuts. In this study, the recalls were conducted nine and two months after the crop cut in 2021 and 2022, respectively. Therefore, the estimated negative effect and its implication is consistent with the less APE in 2021 than 2022 (Fig. 7) and suggestive that farmer recalls get closer to crop cuts if the recall is done after several months (to a year or before new outputs are harvested) prior to harvesting. Further discussion of the time effect on measurement error is provided subsequently.

3.6. Cost analysis of crop cuts and farmer recall surveys

The cost analyses for yield data estimation confirm that yield estimation under crop cuts is significantly higher than estimation through farmer recall surveys (Fig. 8). The cost gap in terms of Ghana cedis between crop cuts and farmer recall is about 49–52%. The decline in the USD denominated cost for both farmer recall surveys and crop cuts in 2022 relative to 2021 is due to the high depreciation of the Ghana cedis in 2022 and early 2023 when the data were collected. This estimated cost gap could be lesser because the recall survey tool (questionnaire) included other sections in addition to those related to yield estimation, which extended the time spent in interviewing a farmer and reduced the number of farmers that could be interviewed daily.

Table 1
Definition and descriptive statistics of production and socioeconomic factors of the farmers.

Variable	Definition/ measurement	Mean	Std. dev.	Min	Max
Crop cut yield	Maize yield estimated through crop cut (kg/ ha)	2593.9	864.6	627.6	5178.1
Recall yield	Maize yield estimated through farmer recall (kg/ha)	1544.6	803.9	617.8	4942.0
MAE	Mean absolute error (kg/ha)	1049.3	930.3	−1973.5	3742.6
APE	Absolute percentage error (%)	36.4	32.7	−118.7	81.3
Farm size	Farmer recall of the total area cultivated with maize (ha)	1.9	1.6	0.2	10.5
Labor	Number of laborers used (#/ha)	7.7	4.1	1.0	34.6
Education	Number of years of formal education (years)	4.8	5.6	0.0	16.0
Experience	Number of years of involvement in maize production (years)	12.9	10.4	1.0	60.0
Age	Actual years from birth of the farmer to the time of data collection (years)	39.4	11.5	19	75
FBO	Dummy: 1 if the farmer belongs to a farmer group and 0 if not	0.35	0.5	0.0	1.0
Extension	Dummy: 1 if the farmer had access to extension services and 0 if not	0.54	0.5	0.0	1.0
Farming system	Dummy: 1 if the farmer engaged in a farming system other than monocropping and 0 if monocropping	0.89	0.32	0	1
Fertilizer use	Dummy: 1 if the farmer used any form of mineral fertilizers for maize production and 0 if not	0.72	0.45	0	1
Credit access	Dummy: 1 if the farmer had access to agricultural credit and 0 if not	0.40	0.49	0	1
Non-farm activity	Dummy: 1 if the farmer engaged in other economic activities in addition to farming and 0 if not	0.51	0.50	0	1

Farmer recall surveys have generally been noted to be cheap, quick, and less laborious and, hence, are a useful method for estimating yields on a larger scale (Sapkota et al., 2016). However, the potential underestimation of yields under this method means that there will be excessive resource allocations towards improving maize production in the country. For instance, given the low average recall yields (Table 1) and Ghana's maize potential yield of 5.5 t/ha, the yield gap is as high as 4 t/ha while this gap is only 2.9 t/ha if the average crop cut yield of 2.6 t/ha is compared with the potential yield. Closing the estimated 4 t/ha yield gap will require about double the financial, capital, and technological investments by various agricultural development agencies, donors and government of Ghana if the gap is realistically 2.9 t/ha. The implication is that beyond the differences in the cost required to generate yield data under various methods, the implications for budgetary allocation to the agriculture sector and design of appropriate policy could be dire.

Table 2
Effect of cultivated area on absolute percentage error.

Measurement	2021			2022		
	Coef.	Std. Err.	*P*-value	Coef.	Std. Err.	P-value
Model 1						
Farm size (ha)	3.454**	1.394	0.014	1.820	1.298	0.162
Constant	27.059	3.907	0.000	34.446	2.901	0.000
Model 2 *First quantile is used as a reference group*						
Second quantile	18.558***	6.676	0.006	7.048	4.374	0.108
Third quantile	20.772***	6.766	0.002	9.353*	5.153	0.070
Fourth quantile	22.865***	7.587	0.003	9.508*	5.123	0.064
Constant	20.969	4.309	0.000	32.142	3.031	0.000

Note: The dependent variable is APE. ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

3.7. Explaining the sources of systematic measurement error in yields

The data show a significant difference in the yield estimates, with crop cuts having higher yield estimates than those from farmer recall (Fig. 3). These differences can be explained as systematic bias and sampling error. Several interrelated systematic biases are plausible explanations. Because crop yield is estimated using the measured/reported output and cultivated area, it thus suggests that the systematic biases identified were results of misreporting of either the output, cultivated area or both.

A potential source of yield measurement error is a deliberate underreporting of harvested output or cultivated area by the farmers. Farmers may be incentivized to underreport their output as a way of registering their vulnerability and need for targeting under an intervention or program. This is especially so since the enumerators introduced themselves as reps from an international agricultural research organization, which has implemented an intervention in some of the farmers' communities. This is not to blame farmers for inaccurate reporting on their production figures; it indicates the need for providing a detailed explanation to farmers on the objectives of recall surveys. Gourlay et al. (2019) and Sapkota et al. (2016) also suggested that farmers may underreport their yields to indicate their suitability for programs or services. In some circumstances also, it becomes clear during interviews that some respondents provide inconsistent information when quizzed further or alternatively. For instance, the reported total output data for some farmers when asked 'what is the total output obtained' do not equal to the sum of when they were asked the specific quantities for the various output distributions paths (i.e. direct home consumption, sale, gifting, post-harvest loss). For cultivated area misreporting for instance, some farmers cultivated multiple plots of maize and therefore had to recall the specific size of these multiple sizes for summation as the farm size of the farmer. This often results in misreporting due to the inability of the farmers to precisely report the size of each farm. The lack of feedback to farmers by previous researchers also influences respondents' attitude. For instance, during the surveys, some farmers expressed their survey participation fatigue by indicating that 'we have been providing our information to you always and afterwards, we do not hear anything again'. This lack or seemingly lack of feedback suggests to the farmers that they do not derive any benefit from the information they provide during surveys.

Underreporting of outputs can also be related to gifting the harvested output and using output for in-kind payments. About 15% of farmers who accessed credit repaid this using their harvested maize grains, mostly immediately after harvest. Some of these farmers failed to include the quantity of their harvest used for loan repayment in their total output when recalling. Similarly, some farmers fail to include post-harvest losses in total yields. Such post-harvest losses occur at harvest, during transportation, and during storage. Considering the high post-

Table 3
Factors influencing the measurement error among farmers.

Variable	Logit model				OLS random effect model		
	Coef.	Std. Err.	P > z	Marginal effect	Coef.	Std. Err.	P > z
Quantile							
Second	1.244***	0.454	0.006	0.104	17.447***	3.698	0.000
Third	2.091***	0.591	0.000	0.131	21.431***	4.285	0.000
Fourth	1.400**	0.577	0.015	0.111	19.296***	4.665	0.000
Region							
North East	−1.230*	0.643	0.056	−0.054	−9.107**	4.577	0.047
Savannah	0.219	0.976	0.822	0.005	10.154	6.298	0.107
Upper East	0.317	1.062	0.765	0.007	14.341**	6.921	0.038
Upper West	−2.935***	0.611	0.000	−0.296	−32.583***	3.879	0.000
Other covariates							
Age	0.041**	0.020	0.045	0.002	0.392**	0.151	0.010
Gender	−0.593	0.477	0.213	−0.034	−5.078	4.264	0.234
Education	0.062**	0.033	0.061	0.004	0.440*	0.263	0.094
Experience	−0.061**	0.021	0.004	−0.004	−0.508***	0.168	0.002
FBO	−0.210	0.358	0.557	−0.012	−2.693	3.140	0.391
Extension	−0.203	0.357	0.571	−0.012	−1.317	3.137	0.675
Credit access	−0.081	0.350	0.817	−0.005	4.035	2.955	0.172
Non-farm	−0.444	0.367	0.227	−0.026	−2.006	3.016	0.506
Labor	−0.001	0.043	0.975	0.000	0.130	0.358	0.716
Year	−0.433	0.391	0.268	−0.025	−5.090	3.359	0.130
Fertilizer use	0.018	0.654	0.978	0.001	3.035	4.557	0.505
Constant	3.200	1.358	0.018		31.203	10.382	0.003
lnsig2u	−14.146	117.906			3.599		
Sigma_u	0.001	0.050			28.617		
rho	2.18e-07	2.6E-05			0.016		
Wald chi sq.	47.15***						
Adj. R. sq.					0.2558		

***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

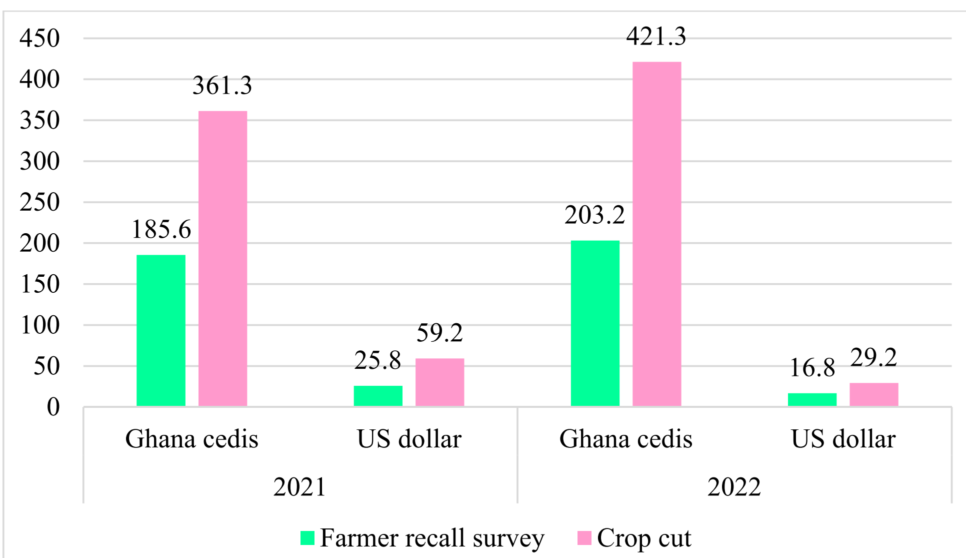


Fig. 8. Estimated cost of gathering a single estimate under farmer recall survey and crop cut methods. The conversion factor of U.S. dollars to Ghana cedis used was 6.0998, 7.20064, 14.4068, and 12.0769 in November 2021 (for 2021 crop cut), March 2022 (for 2021 recall), November 2022 (for 2022 crop cut), and January 2023 (for 2022 recall), respectively. Source. https://www.oanda.com/currency-converter

harvest losses estimated at 5–70% in Ghana (Darfour and Rosentrater, 2016), such exclusion can result in significant yield underestimation. This explains why farmer recall estimates are best described as an estimate of economic yield and not biological yield, as under crop cut (Fermont and Benson, 2011).

As observed in other studies (Gourlay et al., 2019), the sampled farmers and, in some cases, the research assistants tended to approximate the harvested output and cultivated area figures to whole, or round, numbers. The impact of such data approximation becomes very significant when the cultivated area is approximated to the upper round number and the output is approximated to the lower round number. For instance, if a farmer cultivated an area of 1.6 ha and had an output of 3950 kg of grain, the farmer obtained a yield of 2468.8 kg/ha. During the interview, if the farmer conveniently approximates the cultivated area to 2 ha and the output to 4000 kg, the yield becomes 2000 kg/ha, an underestimation of 468.8 kg/ha simply due to rounding. Such approximations are very common among farmers being interviewed, and the compounding effect is the underestimated yield. A further sensitivity analysis due to recall errors on farm size or output is provided in Fig. 9.

The period from harvest to recall survey and the duration of the

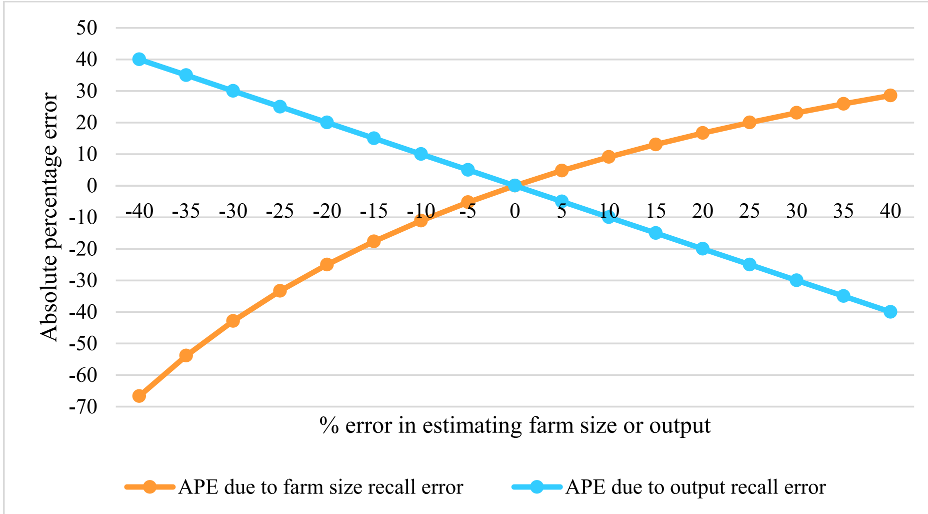


Fig. 9. Sensitivity analysis of APE due to recall error on farm size and output.

recall interviews can also explain the systematic measurement error in the recall surveys. The ability and willingness of farmers to provide factual responses begins to dwindle if the duration of data collection is longer. For this study, data on crop outputs and cultivated area were often asked in the middle of the questionnaire, which took about 45 min to administer. This may affect the responses obtained; a quantitative analysis of the effect of the length of survey interview on measurement error is however required. A longer length of time between harvest and the recall survey generally results in less accurate recollections. However, some farmers do not immediately bag their grains and thus may not be able to provide the actual output until they have been bagged or distributed. This makes such an error farmer specific. Although statistically indifferent, this study established that the APE was lower in 2021, when the recall survey was done about nine months after harvest, than in 2022, when the recall survey was performed about two months after harvest. The implication is that near accurate recall data cannot be gathered only a few months after harvesting; instead, a reasonable time should be allowed. Wollburg et al. (2021) reported that farmers in Malawi and Tanzania overestimated their recall yields relative to full plot cuts when the recall length was longer. However, further analysis of the timing of farmer recall surveys is also necessary to identify the right time to conduct the surveys.

Another systematic bias leading to measurement error is the level of accuracy and precision of the research assistants who gather the recall data. Although the research assistants in this study were trained on the data collection tools and protocols, there were inherent and potential biases. For instance, if the research assistants wanted to finish their allocated daily data entry, it is plausible they may not have allowed ample time for the farmers to reflect on their figures. Research assistants may have also wrongly entered production figures, particularly the cultivated area and the total harvest. For instance, in typing a figure on a mobile app, an adjacent figure can mistakenly be pressed. Some of these errors are unidentifiable during field and central editing and, hence, may have an effect on the estimated yields. Depending on the error, it could result in an underestimation or overestimation. Some outlier figures in particular can be as a result of this error source or from the part of the farmers. This interviewer effect on data measurement has been widely reported (Rashighi and Harris, 2017; West and Blom, 2017; Olson and Peytchev, 2007). Olson and Peytchev (2007), for instance, explained that interview time decreases with the experience of an enumerator, while West and Blom (2017) argued that, in addition to sociodemographic characteristics of enumerators, factors such as the speed and probing skills of the enumerator influence survey errors. This study also found that the recall yield differed by enumerator.

In addition to the potential for underestimation during farmer recall surveys, crop cut yields also have the potential to be overestimated, a common occurrence (Sapkota et al., 2016). One of the potential sources of overestimation in this study was the high heterogeneity of farmers' fields. From land preparation to planting and farm management, the farmers introduced heterogeneity of the farms. For instance, the failure of farmers to plant using recommended spacing meant that the plant population was unevenly distributed. There were instances in which farmers planted improved seeds and applied fertilizers or controlled weeds on only some portions of the farm in a timely manner, thereby resulting in a high heterogeneity in the crop performance on the fields. Therefore, extrapolating crop cuts from only three sections of 4 m^2 each on such farms could potentially result in systematic overestimation. Related to the heterogeneity effect is the non-random selection of subplots for crop cuts. The three sections of each farm that were harvested were discretionarily determined through visual inspection by the yield cut technicians. Therefore, the tendency to select farm areas with relatively high crop stands or potential high-yield areas is high. Evidently, despite the high heterogeneity on the farms, the mean difference between the cuts from the three sections of each farm was only 56–212 kg/ha ±10–39 kg/ha.

Another source of potential overestimation in a crop cut is the total harvest and border effect. In crop cuts, the sections are completely harvested, unlike in reality, where farmers sometimes inadvertently leave some maize stands unharvested, e.g., because the plant has been lodged or unsighted. Although the effect on small farms would be minor, the cumulative effect become significant with larger farms. Also, the tendency to include crops on the border of the cut into the cut area is high. This may result in high crop yields in the cut areas, and extrapolating such values results in an overestimation of the yield. This potential overestimation in crop cuts was confirmed since higher yields under crop cuts and larger farm sizes were associated with high measurement error (Table 4; see Appendix 3 for distribution). Empirical evidence suggests a high potential for higher measurement error on higher yielding plots than the low-yielding plots (Kosmowski et al., 2021).

Table 4
Pairwise correlation between MAE, farm size, crop cut yield, and farmer recall yield.

	MAE	Farm size	Crop cut
Farm size	0.165*		
Crop cut	0.601*	0.160*	
Farmer recall	−0.511*	−0.019	0.380*

* Indicates significant correlation coefficient.

3.8. Sensitivity analysis of APE due to recall error on farm size and output

Farmer recall yields are estimated by dividing the recalled output by the recalled farm size. Therefore, the recall error in any of these values can result in yield measurement error between crop cuts and recall surveys. Thus, a sensitivity analysis of the impact of recall error from either farm size or output was done to determine how the APE would vary in response to such recall errors (Fig. 9). Assuming each farmer cultivated an average of 1.9 ha from the sample and obtained an output of 4928.4 kg of maize, then the yield for each of these farmers would be 2593.9 kg/ha. If each of these farmers made a cumulative error of ±5% in recalling the 1.9 ha, but correctly recalled their output as 4928.4 kg, then their average yield would vary, increasing with an increasing underreporting of farm size and decreasing with an overreporting of the farm size. If the crop cuts of the farmers are accurately measured with a yield of 2592.9 kg/ha, then the APE would be positively rising. On the other hand, when farm size is accurately recalled and the output is reported with error, again with the same 2592.9 kg/ha crop cut yield for each farmer, the APE varies downwardly. Where both the recalled farm size and output are correctly measured, the farmer recall yield equals the crop cut yield and the APE is at zero. Detailed results on the sensitivity analysis are presented in Appendix 4.

4. Conclusions and policy recommendations

The relevance of agricultural data in policy design for agricultural sector transformation and overall sustainable development is undisputable. The lack of accurate data can derail efforts in reducing food insecurity and poverty due to faulty policies informed by faulty data. For a developing country such as Ghana with high potential for the agricultural sector to catalyze development, more attention must be given to obtaining accurate agricultural statistics, such as inputs, outputs, and outcomes. This study used empirical field data to estimate the maize yield measurement error. Farmer recall surveys and crop cuts were used to gather data on maize production in the Guinea Savannah zone, where there is huge potential for increasing maize yield. The results showed a high measurement gap, ranging from −118.7% to 81.3% and averaging 36.4%, between crop cuts and farmer recall yield estimates. This measurement error is due to several systematic biases that either influenced the recall yield or the crop cut yield. However, considering the potential error with both yield estimation methods but the near accuracy of crop cut yield estimation over recall yield, notwithstanding its high budgetary requirements, the crop cut method, rather than farmer recall, should be relied upon in estimating yields. But where economic yields are the topmost priority, farmer recall surveys can be used under appropriate research designs to minimize the identified potential sources of measurement error. Yield estimation involved total harvest and cultivated areas, and either or both parameters can be misreported by the farmer. Although not tested in this study, the use of GPS has proven to be more precise than farmer recalls in estimating the cultivated area. Therefore, recall surveys should consider integrating the use of GPS estimate of the cultivated area in their design to at least offset the error from measuring the cultivated area. The systematic error in yield measurement is also influenced by other socioeconomic characteristics of the farmers. Therefore, farmers should be educated on record keeping in order to avoid memory recalling. Also, there is the need for policy makers, including Ministry of Local Government and Rural Development, Ghana Statistical Service, and MoFA, to sensitize farmers to understand the relevance and the derived benefits they obtain from the information they provide to researchers by participating in surveys.

Interpretation of farm data should also be properly categorized within the given context. The general description of every estimate as yield makes it difficult to compare such estimates with one another; rather, these should be described appropriately either as economic or biological yield. More precision with much attention to potential sources of systematic bias is needed for each yield estimation method to ensure that the yield estimates under any method can be appropriately interpreted. Also, there is a high potential impact of yield measurement errors in policy design, and resource allocation and investment into the agriculture sector. Policy-relevant empirical analyses involve appropriate research designs to guarantee positive outcomes from research recommendations. Further analysis of how the measurement error affects outcomes, such as food security and poverty eradication, both at the micro and macro levels, in Ghana is required. Despite the detailed analyses and relevance of this study, there are two limitations. First, measurement error in both farm size and output can affect yield estimation, hence, the impact on the APE varies. For this study, farm size was measured through recalls only. Therefore, although this study provided evidence on the yield measurement error, it fails to attribute the specific proportions of the error that is due to inaccurate measurement of the farm size or the production/output. This limitation can be addressed by comparing the APE between farm size under farmer recalls and other methods such as GPS estimation method, and their implications on yield measurement. Secondly, a limitation of the farmer recall survey for this study was that the recall data were obtained for a broader objective beyond this study. This led to longer data collection periods. For a methodological comparison, as in this study, the effect of these known systematic biases should be intentionally minimized. Future studies should therefore ensure an appropriate duration for farmer recall data collection.

CRediT authorship contribution statement

William Adzawla: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Edinam D. Setsoafia:** Writing – original draft, Writing – review & editing. **Eugene D. Setsoafia:** Writing – original draft, Writing – review & editing. **Solomon Amoabeng-Nimako:** Writing – original draft, Writing – review & editing. **Williams K. Atakora:** Project administration, Resources. **Prem D. Bindraban:** Conceptualization, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We appreciate Julie Kohler of IFDC for English editing the manuscript. This manuscript is produced using data generated under FERARI program. Our appreciation therefore to Mohammed VI Polytechnic University and the OCP group for the financial support to the FERARI program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.agsy.2023.103817.

References

Adzawla, W., Atakora, W.K., Kissiedu, I.N., Martey, E., Etwire, P.M., Gouzaye, A., Bindraban, P.S., 2021. Characterization of farmers and the effect of fertilization on maize yields in the Guinea Savannah, Sudan Savannah, and transitional agroecological zones of Ghana. EFB Bioecon. J. 1 (100019) https://doi.org/10.1016/j.bioeco.2021.100019.

Adzawla, W., Bindraban, P.S., Atakora, W.K., Camara, O., Gouzaye, A., 2022. Economic viability of smallholder agriculture in the savannah and transitional zones of Ghana: implications of farm output commercialization and farm diversification. Sustaianability 14 (11548), 1–22. https://doi.org/10.3390/su141811548.
Ambler, K., Herskowitz, S., Maredia, M.K., 2021. Are we done yet? Response fatigue and rural livelihoods. J. Dev. Econ. 153 (102736) https://doi.org/10.1016/j.jdeveco.2021.102736.
Beegle, K., Carletto, C., Himelein, K., 2012. Reliability of recall in agricultural data. J. Dev. Econ. 98 (1), 34–41. https://doi.org/10.1016/j.jdeveco.2011.09.005.
Blandford, D., 2007. Information deficiencies in agricultural policy design, implementation and monitoring (OECD Food, Agriculture and Fisheries Working Papers, No. 6. OECD Publishing. https://doi.org/10.1787/067228574571.
Boullouz, M., Bindraban, P.S., Kissiedu, I.N., Kouame, A.K.K., Devkota, K.P., Atakora, W. K., 2022. An integrative approach based on crop modeling and geospatial and statistical analysis to quantify and explain the maize (Zea mays) yield gap in Ghana. Front. Soil Sci. 2, 1–20. https://doi.org/10.3389/fsoil.2022.1037222.
Carletto, C., Jolliffe, D., Banerjee, R., 2015. From tragedy to renaissance: improving agricultural data for better policies. J. Dev. Stud. 51 (2), 133–148. https://doi.org/10.1080/00220388.2014.968140.
Darfour, B., Rosentrater, K.A., 2016. Maize in Ghana: an overview of cultivation to processing. ASABE Annu. Int. Meet. Pap. 2016, 1–16. https://doi.org/10.13031/aim.20162460492.
Desiere, S., Jolliffe, D., 2018. Land productivity and plot size: is measurement error driving the inverse relationship? J. Dev. Econ. 130, 84–98. https://doi.org/10.1016/j.jdeveco.2017.10.002.
Fermont, A., Benson, T., 2011. Estimating yield of food crops grown by smallholder farmers: A Review in the Uganda Context. In *IFPRI Discussion Paper* (No. 01097). http://cdm15738.contentdm.oclc.org/utils/getfile/collection/p15738coll2/id/124924/filename/124925.pdf.
Giller, K.E., 2020. The food security conundrum of sub-Saharan Africa. Glob. Food Sec. 26, 100431 https://doi.org/10.1016/j.gfs.2020.100431.
Gourlay, S., Kilic, T., Lobell, D.B., 2019. A new spin on an old debate: errors in farmer-reported production and their implications for inverse scale - productivity relationship in Uganda. J. Dev. Econ. 141 (August 2017), 102376 https://doi.org/10.1016/j.jdeveco.2019.102376.
GSS, 2019. Ghana Living Standards Survey round 7 (GLSS7): Main report. https://statsghana.gov.gh/gsspublications.php?category=MTAwMjg3Mzk3NC4zMDc=/webstats/1opr93rn57.
GSS, 2019a. 2017/18 Ghana census of agriculture: National report. https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/FinalReport11112020printedversion.pdf.
GSS, 2023. Sectorial share of GDP. https://www.statsghana.gov.gh/.
Kosmowski, F., Chamberlin, J., Ayalew, H., Sida, T., Abay, K., Craufurd, P., 2021. How accurate are yield estimates from crop cuts? Evidence from smallholder maize farms in Ethiopia. Food Policy 102, 102122. https://doi.org/10.1016/j.foodpol.2021.102122.
Kouame, A.K.K., Bindraban, P.S., Kissiedu, I.N., Atakora, W.K., El Mejahed, K., 2023. Identifying drivers for variability in maize (Zea mays L.) yield in Ghana: a meta-regression approach. Agric. Syst. 209 (April 2022), 103667 https://doi.org/10.1016/j.agsy.2023.103667.
Lauer, J., 2002. Methods for calculating corn yield. In: Agronomy Advice. https://green.extension.wisc.edu/files/2010/05/Corn-Yields.pdf.
Lobell, D.B., Azzari, G., Burke, M., Gourlay, S., Jin, Z., Kilic, T., Murray, S., 2019. Eyes in the sky, boots on the ground: assessing satellite- and ground-based approaches to crop yield measurement and analysis. Am. J. Agric. Econ. 102 (1), 202–219. https://doi.org/10.1093/ajae/aaz051.
MoFA, 2021. Facts and figures: Agriculture in Ghana 2020.
Nature, 2020. To end hunger, science must change its focus. Nature 586, 336. https://doi.org/10.1038/d41586-020-02849-6.
Nature Plants, 2020. Feast and famine in agricultural research. Nat. Plants 6 (10), 1195. https://doi.org/10.1038/s41477-020-00795-9.
Olson, K., Peytchev, A., 2007. Effect of interviewer experience on interview pace and interviewer attitudes. Publ. Opin. Q. 71 (2), 273–286. https://doi.org/10.1093/poq/nfm007.
Paliwal, A., Jain, M., 2020. The accuracy of self-reported crop yield estimates and their ability to train remote sensing algorithms. Front. Sustain. Food Syst. 4 (March), 1–10. https://doi.org/10.3389/fsufs.2020.00025.
Rashighi, M., Harris, J.E., 2017. Mode and interviewer effects in egocentric network research. Physiol. Behav. 176 (3), 139–148. https://doi.org/10.1053/j.gastro.2016.08.014.CagY.
Sapkota, T.B., Jat, M.L., Jat, R.K., Kapoor, P., Stirling, C., 2016. Yield estimation of food and non-food crops in smallholder production systems. In: Rosenstock, T.S., Rufino, M.C., Butterbach-Bahl, K., Wollenberg, E., Richards, M. (Eds.), Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture, pp. 163–174. https://doi.org/10.1007/978-3-319-29794-1.
Tandzi, L.N., Mutengwa, C.S., 2020. Estimation of maize (Zea mays L.) yield per harvest area: appropriate methods. Agronomy 10 (1), 1–18. https://doi.org/10.3390/agronomy10010029.
Tetteh, F., Larbi, A., Nketia, K.A., Senayah, J.K., Hoeschle-Zeledon, I., Abdul-Rahman, N., 2016. Suitability of soils for cereal cropping in northern Ghana. International Institute of Tropical Agriculture. www.africa-rising.net.
Wahab, I., 2020. In-season plot area loss and implications for yield estimation in smallholder rainfed farming systems at the village level in sub-Saharan Africa. GeoJournal 85 (6), 1553–1572. https://doi.org/10.1007/s10708-019-10039-9.
West, B.T., Blom, A.G., 2017. Explaining interviewer effects: a research synthesis. J. Surv. Stat. Methodol. 5 (2), 175–211. https://doi.org/10.1093/jssam/smw024.
Wineman, A., Anderson, C.L., Reynolds, T.W., Biscaye, P., 2019. Methods of crop yield measurement on multi-cropped plots: examples from Tanzania. Food Secur. 11 (6), 1257–1273. https://doi.org/10.1007/s12571-019-00980-5.
Wollburg, P., Tiberti, M., Zezza, A., 2021. Recall length and measurement error in agricultural surveys. Food Policy 100, 102003. https://doi.org/10.1016/j.foodpol.2020.102003.