Disentangling agronomic and economic yield gaps in Ethiopian wheat based systems for better targeting of development interventions (Yield Gap Wheat Ethiopia)

Report #2: Data availability, description and quality

Date: 12th December 2018

Authors: João Vasco Silva, Tom Morley, Marloes van Loon,
Pytrik Reidsma and Martin K. van Ittersum
Wageningen University & Research

Collaborators: Kindie Tesfaye, Moti Jaleta, Frédéric Baudron
International Maize and Wheat Improvement Centre

1) Background

Different datasets are available to estimate and explain wheat yield gaps in Ethiopia and these were briefly summarized in the 1st deliverable of the project. In short, these include 1) spatially explicit water-limited and actual yields for wheat in Ethiopia during the period 2001 – 2012 from the Global Yield Gap Atlas (GYGA) and 2) household surveys from diverse sources containing detailed information on crop yield, management practices and socio-economic conditions for a large of number of farms in the main wheat producing areas of Ethiopia. The latter include the Living Standards Measurement Survey (LSMS) collected by the World Bank; Wheat Adoption and Impact Surveys (WAIS) collected by CIMMYT and; detailed surveys on labour use also collected by CIMMYT within the FACASI project. In addition, for a more detailed regional analysis, we plan to use for one region (probably Arsi), variety trial data for different elevations.

The main purpose of this deliverable is two-fold: 1) to detail which data will be used for a national and regional yield gap analysis in the next six months of the project and 2) to provide some preliminary and descriptive results on wheat production in Ethiopia based on two household surveys. We will also highlight data quality issues we came across so far.

2) Datasets available and progress

- **Global Yield Gap Atlas**

GYGA is an important source of biophysical and agronomy data for yield gap analysis. Background information on the simulation of water-limited yields for cereals (including wheat) in Ethiopia can be found in http://www.yieldgap.org/en/web/guest/ethiopia. Potential yields (Yp) and water-limited yields (Yw) are available per weather station / climate zone for the period 1998 – 2011 only. Currently, we are updating the yield gaps of wheat in Ethiopia for the period 2011 – 2017. To do this, the protocol of the GYGA (www.yieldgap.org; Grassini et al., 2015; van Bussel et al., 2015) is followed. The protocol is based on a bottom-up approach, in which Yp and Yw are estimated in a spatial framework with the crop model WOFOST based on local weather, soil and agronomic data.

Daily weather data from weather stations in Ethiopia were acquired from the National Meteorology Agency of Ethiopia in close cooperation with dr Kindie Tesfaye. The estimation of wheat yield gaps will be done for 12 weather stations, which were selected based on the cultivated wheat area surrounding each weather station. All these weather stations had missing data and we will be able to use observed weather data from 8 weather stations only; for the other 4 weather stations we will use gridded weather data instead given the large amount of missing data. From the 8 weather stations which will be used, around 20% of the temperature and rainfall records were missing. Missing temperature records will be generated using the relation between NASA-POWER and observed data (as described in van Wart et al., 2015), while missing rainfall data will be filled based on data from the Tropical Rainfall Measuring Mission (TRMM) of NASA. Once data cleaning is finalized, we will be able to run the crop model WOFOST and estimate Yp and Yw of wheat for the more recent years. Recent data for actual yields (Ya) have been obtained from the Central Statistical Authority (CSA).

- **LSMS Household Survey**

The LSMS is a nationally representative panel dataset, which was conducted in three waves: 2011, 2013 and 2015. Crop cut yields are available for all three waves but significantly higher wheat yields were observed in 2011 possibly due to a change in survey protocol. Therefore, we will conduct the analysis for all years and for 2013 and 2015 only to see how the 2011 affects the results obtained in the other years.

Our initial plan was to use the LSMS for a national yield gap analysis given the national representativeness of these data and our experiences with these in previous projects. However, the unreliability of crop areas based on farmer recall (see Appendix A) forced us to 1) use crop cut yield data which reduced the sample size available for statistical analyses considerably and 2) refrain us from using continuous variables on input use (e.g. N applied in kg ha^{-1}). For these reasons, we find it hard to build a yield gap analysis at national level solely using the LSMS data. Instead, we propose to use these data for more focused analysis (e.g. to assess the effect of crop establishment method on crop yields) and/or as a secondary source to assess the uncertainty of the results obtained with WAIS dataset. We expect the latter will receive lower priority than the former and its execution will depend on the progress of the project. Descriptive statistics and preliminary results from this data source are provided below.

- **WAIS Household Survey**

The WAIS survey is a panel of households and was conducted in two rounds covering the growing seasons of 2011 and 2014. These data have been used for different purposes in previous publications by Shiferaw et al. (2014) and Abro et al. (2017, 2018) and its large sample size and national coverage makes them highly suitable for the purpose of this project.

The large sample size and national representativeness of the WAIS data make them suitable for yield gap analysis at national and regional level. Previously, we thought of using these data for analysis at regional level only but the first explorations of the data revealed potential to go beyond that (see below descriptive statistics and some preliminary results). Therefore, we propose to conduct statistical analyses both for the regions in which sample sizes are largest (e.g., Arsi, Gojam, Shoa and Tigray) and for the pooled sample (national level) so that results between these can be compared. In addition, we propose to zoom-in on one region (probably Arsi, Figure 3, given its large potential for wheat cultivation) where more detailed analysis will be conducted at crop and farm level. For Arsi, we will be able to supplement the farm data with variety trials conducted at different elevations in this region, as summarized by Bezabih et al. (2018). This adds an additional layer to the analysis which we consider important to explore.

- **FACASI Household Survey**

The FACASI survey was conducted in 2012 with the purpose of mapping the potential demand for mechanisation in Eastern and Southern Africa. These data were used by Silva et al. (under review) to understand the role of labour in explaining yield gaps in Hawassa and Asella, Southern Ethiopia.

The strength of this dataset is the level of detailed information on labour use at crop and farm level. Previous analysis of these data indicated yield gaps are as large as 80% of Yw and that there is competition for labour between wheat and other crops in key periods of the growing season (Silva et al., under review). Some interesting aspects that could be explored further with these data are 1) the use of TEDs to generalize the findings observed in Asella to other areas in the Ethiopian highlands or 2) to analyse labour allocation to different crops at farm level. Both options are particularly useful to map options for mechanization across the region but not necessarily to gain additional insights into the causes of wheat yield gaps in Ethiopia. Therefore, the analysis of these data will receive low priority in the context of this project.

3) Descriptive statistics

3.1) LSMS-ISA Household Survey

Table 1 contains descriptive statistics from the **LSMS-ISA** survey for plots on which wheat was crop-cut in the 2013 and 2015 Meher seasons. The 2015 (wave 3) survey marked a significant expansion in the number of crop cut wheat fields as compared to 2013 (wave 2), with most of the expansion occurring in Amhara and Oromia.

The **crop cup wheat yield** has remained relatively stable across all regions and seasons with yields averaging between 1.0 and 1.5 t FM ha^{-1}. One notable exception is Oromia in the 2015 Meher season in which the crop cut yield appears to have been above 2.0 t FM ha^{-1}. As part of this project we will investigate the reasons for this increase and in particular how it relates to crop damage due to climate and pests as well as the geographical location of the high yielding plots. Figure 2 includes crop cut yields from the first wave of the survey (Meher Season 2011) which we initially chose not to include due to a different crop cutting protocol (crop cut on 4m^2 area as opposed to 16m^2). Tracking the evolution of crop cut yields for the same farmers across time indicates that the high yields observed in 2015 are not out of line with those in 2011. It remains to be seen why the 2013 yield deviates from the 2011 and 2015 yields.

To summarise the remainder of Table 1, we see little use of **irrigation**, or **crop residue** (after planting) in any region, and strong evidence of farmer reported **damage**, which is often due to a lack of rainfall or pests. Only a small proportion of fields are left **fallow** between seasons and

planting tends to be performed using oxen, as opposed to by tractor or by hand. Most farmers report broadcast planting rather than row-planting. In terms of input use, **improved seed** use is more common in Tigray and SNNP than in Amhara or Oromia but uptake of improved seed is low in general and rarely exceeds 15% of wheat plots. **Fertilizers** are widely used, especially in the SNNP and Tigray regions. Fertilizer use exceeds 50% of plots in all regions and both seasons. **Biocides** are applied in all regions, but most frequently in Oromia and SNNP. Most of the biocide use is herbicide with only a small proportion of farmers reporting pesticide or fungicide on their plots in any region or year.

Table 1. Descriptive statistics for wheat production systems across two time periods and four administrative regions in Ethiopia. Yield data were collected using crop cuts and all data refer to the Meher season. ***Source: LSMS-ISA.***

	AMHARA		OROMIYA		SSNP		TIGRAY	
	2013	2015	2013	2015	2013	2015	2013	2015
Observations (n)	81	171	89	148	87	90	34	54
Wheat yield (t FM/ha)	1.29	1.27	1.28	2.13	1.30	1.17	1.30	1.49
Irrigation (1=yes)	0.00	0.01	0.00	0.00	0.00	0.00	0.06	0.04
Crop residues (1=yes)	0.02	0.02	0.01	0.03	0.00	0.03	0.00	0.06
Damage (1 = yes)	0.36	0.44	0.25	0.44	0.28	0.69	0.50	0.89
Fallow (1 = yes)	0.14	0.10	0.10	0.16	0.15	0.08	0.00	0.02
Ox ploughings (#)	3.25	3.37	3.75	3.68	3.20	3.76	3.24	3.07
Ox planting (1 = yes)	0.93	0.88	0.87	0.84	0.70	0.82	0.85	0.76
Manual planting (1 = yes)	0.00	0.01	0.00	0.03	0.23	0.08	0.00	0.00
Tractor planting (1 = yes)	0.00	0.01	0.01	0.01	0.00	0.02	0.00	0.00
Broadcasting (1 = yes)	0.90	0.75	0.89	0.94	0.98	0.80	1.00	0.94
Row planting (1 = yes)	0.10	0.25	0.11	0.06	0.01	0.20	0.00	0.04
Improved variety (1 = yes)	0.11	0.05	0.06	0.03	0.09	0.17	0.21	0.11
Fertiliser use (1 = yes)	0.53	0.63	0.69	0.67	0.63	0.82	0.85	0.76
N applied (1 = yes)	0.53	0.63	0.69	0.67	0.63	0.82	0.85	0.76
P applied (1 = yes)	0.47	0.58	0.69	0.64	0.62	0.81	0.79	0.69
Manure use (1=yes)	0.28	0.15	0.09	0.10	0.14	0.08	0.41	0.52
Biocide use (1 = yes)	0.11	0.22	0.64	0.70	0.39	0.61	0.03	0.11
Herbicide use (1 = yes)	0.11	0.22	0.64	0.68	0.39	0.60	0.03	0.07
Pesticide use (1 = yes)	0.00	0.00	0.01	0.11	0.03	0.01	0.00	0.04
Fungicide use (1 = yes)	0.00	0.01	0.01	0.17	0.02	0.01	0.03	0.00
Oxen owned (#)	1.41	1.66	1.98	2.05	1.12	1.83	1.30	1.28

3.2) WAIS Household Survey

Descriptive statistics of the **WAIS data** are presented only for the administrative zones with more than 100 wheat plots surveyed in both 2011 and 2013 (**Table 2**). Only 5 out of the 25 zones covered by the survey fulfilled this criteria namely Arsi (n = 1078), West Arsi (n = 620), North Shoa (n = 647) and East Shoa (n = 427) in Oromia province and, South Wollo (n = 846) and East Gojam (n = 311) in Amhara province.

Wheat production in the Ethiopian highlands is mostly **rainfed** as in either year irrigated wheat plots were less than 3% of the entire sample. **Wheat yields** across the selected zones were on average 1.78 and 1.84 t DM ha^{-1} in 2011 and 2013, respectively. These are slightly lower, and provide no evidence of yield progress, than the national average of 2.03 and 2.45 t DM ha^{-1} reported in official statistics for the same years. Conversely, there were yield differences between administrative zones as wheat yields were consistently smallest in South Wollo (ca. 1.2 t DM ha^{-1}) and greatest (> 2.0 t DM ha^{-1}) in Arsi, West Arsi and East Shoa in both years.

Land preparation for wheat comprises 4 – 5 **ox ploughings** in all zones except South Wollo, where ca. 3 ox ploughings are more common. This operation takes between 20 – 30 days ha^{-1} across the different zones, being shortest in South Wollo and longest in East Gojam. The management of **crop residues** varies per year: 15 – 33% of the wheat plots surveyed in 2011 had crop residues incorporated which was true for less than 10% of the wheat plots in 2013.

Urea (46.6% N) and di-ammonium phosphate (DAP, 18%N and 20% P) were the only **mineral fertilisers** applied to wheat during the surveyed years. There were no major differences in N and P application rates across years which were on average 43.7 kg N ha^{-1} and 18.0 kg P ha^{-1}, respectively. In both years, N application rates were greatest (50 – 70 kg N ha^{-1}) in East Gojam, East Shoa and North Shoa and lowest (15 – 25 kg N ha^{-1}) in Arsi and West Arsi while there were no major differences in P application rates between zones. **Organic fertilisers** in the form of cattle manure were used in less than 25% of the wheat plots in both years across all zones. The use of manure was particularly low in East Shoa (ca. 7% of the sample).

Pre-emergence herbicides and hand-weeding are the main **weed control** measures used by farmers. On average, herbicides were applied at a rate of 0.73 L/ha and 1.1 hand-weeding was performed. There were no major differences observed across zones, the only exception being South Wollo where herbicides were barely used and up to two hand-weeding operations were performed instead. The use of **pesticides** for pest and disease control is negligible as indicated by the very low application rates across the sample.

Three main sources of **farm power** are used for wheat production namely manual labour, animal draught power and, in specific zones, machinery. The use of manual labour for land preparation, hand-weeding and harvesting was on average 68.5 person-day/ha, there being clear differences between zones. For instance, labour use was nearly half in Arsi and West Arsi (45.6 person-day/ha) compared to the other zones (79.9 person-day/ha). In case of animal draught power, farmers had on average 10.6 topical livestock units (TLU) and owned 1.98 pairs of oxen. In general, farmers had more TLU and oxen pairs owned in zones where labour use was lowest and vice-versa. Hiring labour for wheat cultivation is common across all zones while hiring oxen is most common in Arsi and East Shoa zones. Finally, hiring of machinery is particularly common in West Arsi and Arsi but not as much in the other zones.

Table 2. Mean values of the input-output coefficients for wheat production in the Meher season across the Ethiopian highlands. Data are presented for regions with more than 100 observations in both 2011 and 2013 growing seasons. *Source: CIMMYT WIAS.*

MEAN VALUES	ARSI		WEST ARSI		NORTH SHOA		EAST SHOA		SOUTH WOLLO		EAST GOJAM	
	2011	2013	2011	2013	2011	2013	2011	2013	2011	2013	2011	2013
Observations (n)	568	510	319	301	325	322	225	202	438	408	162	149
Wheat yield (t DM/ha)	1.93	2.16	2.19	2.16	1.41	1.81	2.01	2.12	1.22	1.18	1.90	1.62
Irrigation (1=yes)	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.03	0.00
Crop residues (1=yes)	0.28	0.03	0.34	0.06	0.18	0.02	0.15	0.10	0.19	0.00	0.33	0.03
Ox ploughings (#)	4.79	4.64	4.41	4.25	3.89	3.76	4.54	4.07	3.23	3.04	4.99	4.54
Ox ploughing (days/ha)	26.79	27.52	25.24	23.14	28.55	30.00	23.34	24.51	21.74	19.87	30.23	30.77
N applied (kg N/ha)	16.41	19.37	21.53	22.30	54.77	65.92	55.70	56.74	51.91	41.52	67.06	51.18
P applied (kg P/ha)	13.80	17.88	16.49	18.30	15.01	23.94	21.80	24.94	12.35	14.49	18.56	18.23
Manure use (1=yes)	0.17	0.18	0.10	0.13	0.19	0.17	0.07	0.07	0.22	0.23	0.25	0.09
Hand-weedings (#)	0.99	0.60	0.98	0.87	0.90	0.89	0.97	0.85	1.95	1.80	1.14	1.20
Herbicide use (L/ha)	0.61	0.92	0.68	0.90	3.76	0.32	0.61	0.54	0.00	0.01	0.29	0.12
Pesticide use (L/ha)	0.08	0.04	0.03	0.03	0.04	0.01	0.02	0.01	0.00	0.00	0.01	0.04
Labour use (pers-day/ha)	46.04	53.71	41.09	41.66	69.11	96.06	67.14	53.80	87.98	103.59	79.15	82.68
TLU (-)	11.78	13.56	13.26	13.55	12.64	11.79	11.39	11.10	6.38	6.15	7.44	7.60
Oxen pairs owned (#)	2.37	2.60	2.23	2.53	1.91	1.87	2.41	2.39	1.11	1.06	1.61	1.51
Hired ox (ETB)	3.83	0.96	0.00	0.40	0.62	0.00	7.86	6.19	0.00	0.74	0.00	0.00
Machinery (ETB)	81.29	269.78	220.90	611.03	1.25	0.00	4.04	90.87	0.14	0.00	0.00	0.00
Hired labour (ETB)	20.48	44.26	21.30	34.87	5.90	52.95	127.87	202.03	3.84	10.73	17.77	18.89

4) Preliminary results

4.1) LSMS-ISA Household Survey

Several important points emerged while preparing and investigating the data from the LSMS-ISA survey. First, it quickly became apparent that many quantitative variables lacked consistency when compared to external evidence or when tracked across different waves of the same dataset. In response to this we decided to focus on crop cut yields and binary or categorical variables referring to crop management (please see Appendix A attached to this report for further information). In the course of investigating the crop cut yields, and with reference to the crop cutting protocol, it was decided to include all three waves of the crop cut despite earlier hesitations over the expansion of the crop cut area from 4m^2 in wave 1 to 16m^2 in waves 2 and 3. More details on the crop cutting protocol can be found in Appendix A.

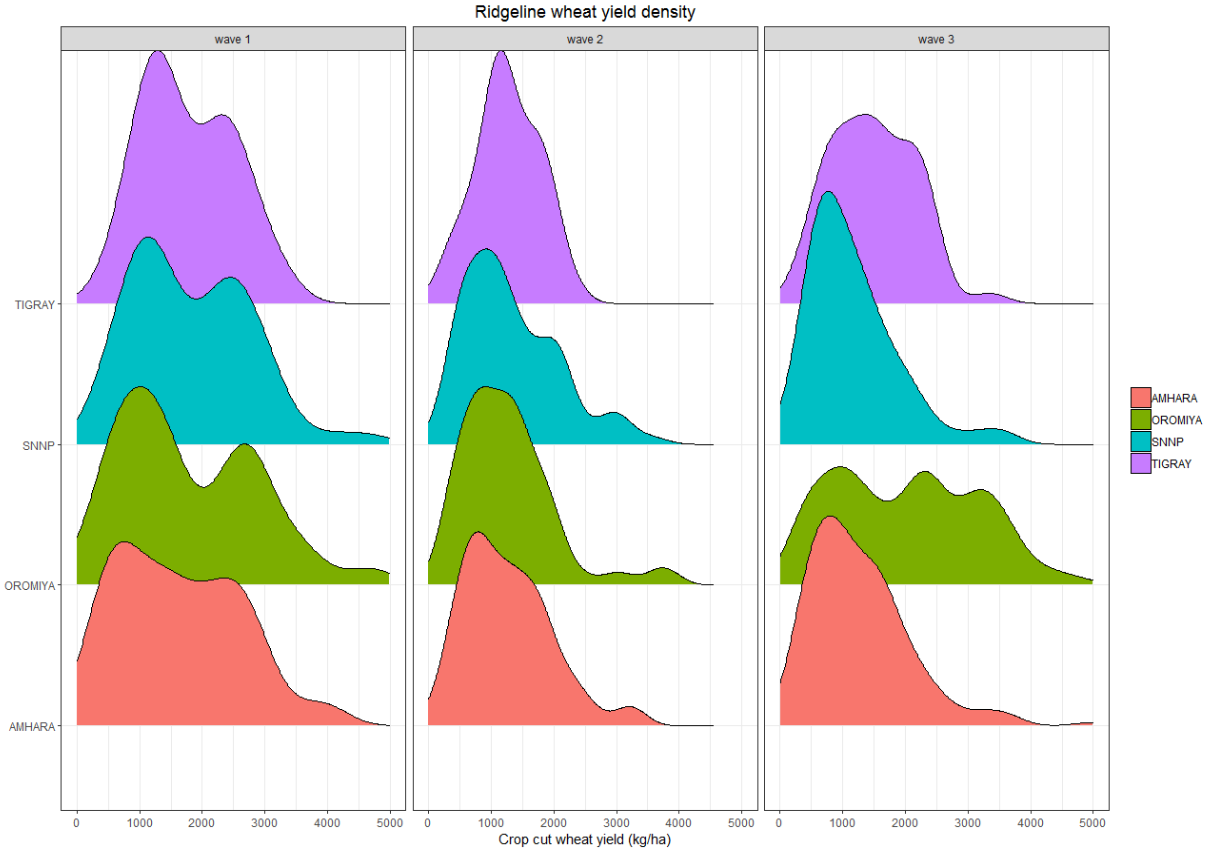

Figure 1. Ridgeline wheat yield density across three time periods and four regions in Ethiopia. Wheat yields were obtained using crop cuts and data refer to the Meher season only. ***Source: LSMS-ISA.***

Figures 1 and 2 present evidence on the wheat yield variation across all three waves of the data. In **Figure 1** the density of the crop cut wheat yields across waves 1, 2, and 3 (2011, 2013 and 2015) Meher seasons are plotted per region. The distribution of the wave 2 wheat yields is markedly narrower than the wave 1 and 3 data, and does not exhibit the same bi- or tri-modal density exhibited by all four regions in wave 1 and in Oromia in wave 3. **Figure 2** shows the evolution of crop cut wheat yields by wave and region. Points connected by a line refer to the same farm at different points in time (but different parcels of land). The decrease in wheat yield in wave 2 in Oromia is notable. When we explore the determinants of the wheat yield (see Section 5), particular attention will be paid to explaining some of these differences over time. The fact that wheat was cultivated by some farms in more than one season, and these can be connected over time, opens the door to using longitudinal analysis techniques.

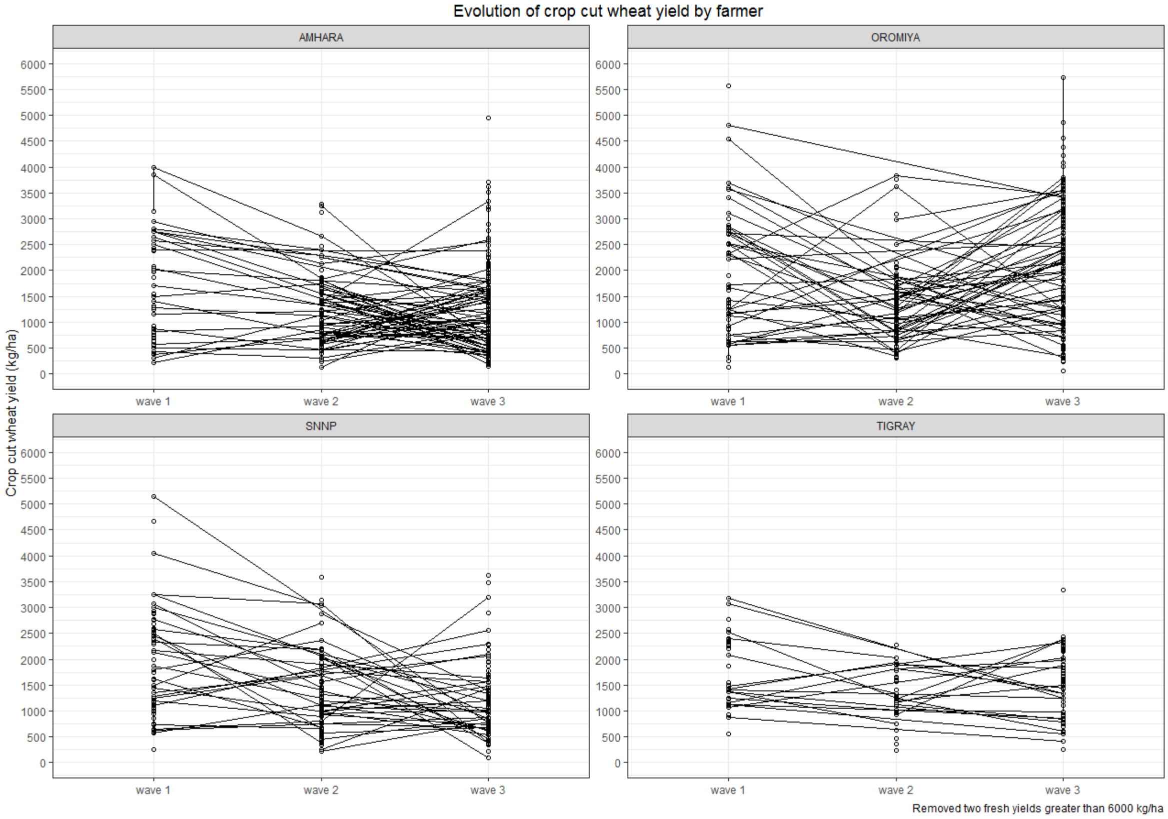

Figure 2. Evolution of crop cut yield data for an individual farm across three time periods and four regions in Ethiopia. Data refer to the Meher season only. ***Source: LSMS-ISA.***

4.2) WAIS Household Survey

The variation in wheat yields and selected inputs, and their relation, are depicted in **Figure 3** for Arsi and West Arsi zones (referred simply as Arsi from here onwards), which are the highest yielding wheat areas of Ethiopia (Table 2). Similar information is available for East and North Shoa and East Gojam in Figures B1 and B2, respectively (see Appendix B), but we refrain to discuss these in detail in this report.

Wheat yields between 6.0 and 8.0 t ha^{-1} were observed in some plots in Arsi both in 2011 and 2013. These relatively high yields are possible from a biophysical perspective as water-limited yields are close to 10.0 t ha^{-1} (www.yieldgap.org). However, yields lower than 1.0 t ha^{-1} were found in many plots, indicating there is a large variation in crop yields between fields and farms and considerable scope to improve wheat production in this area.

Most wheat plots were ox-ploughed between 2 and 9 times (**Figure 3A**). Quantile regressions fitted to the 98th and 90th percentiles indicate a positive linear relation between the number of ploughs and wheat yields. This corroborates the findings of Abro et al. (2018) using the same dataset. However, the actual days used ploughing ranged from a minimum of ca. 10 days ha^{-1} up to more than 50 days ha^{-1} in some plots (**Figure 3B**). This large variation may be explained

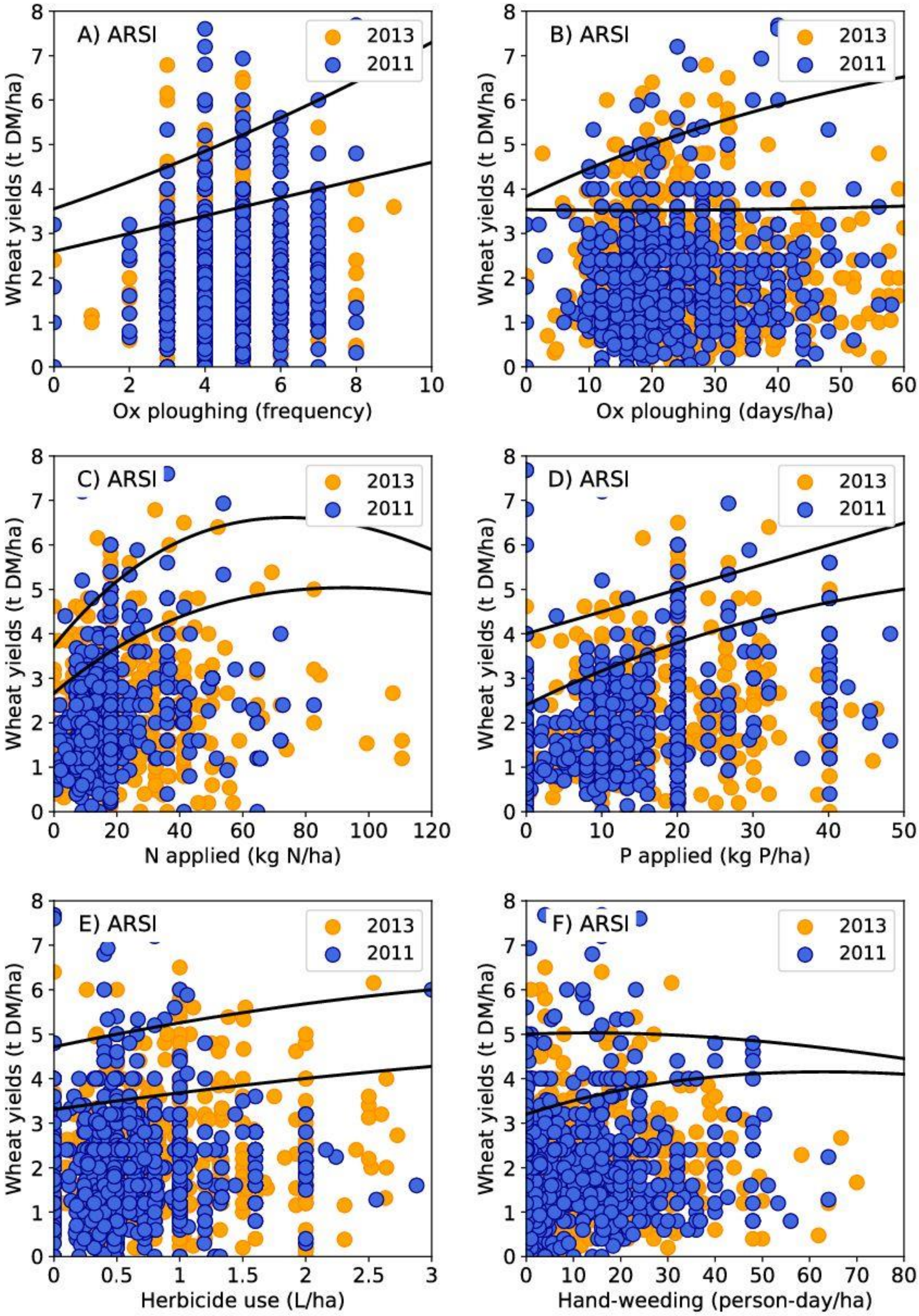

Figure 3. Wheat yield responses to A) ox ploughing frequency, B) ox ploughing days, C) N applied, D) P applied, E) herbicide use and F) hand-weeding in Arsi and West Arsi zones, Ethiopia. Data are presented for the Meher season only. Solid lines show fitted quantile regressions ($y = a + b \times 0.99^{x} + c \times x$) to the 98^{th} (top) and 90^{th} (bottom) percentile. The 98^{th} percentile capture the influence of extreme observations while the 90^{th} percentile refers to the lower threshold for highest farmers' yields commonly used.

by the number of oxen pairs owned by each household but this remains to be tested. Wheat yields were positively associated with the number of days used for ploughing but only when quantile regressions were fitted to the 98th percentile.

N application rates were smaller than 60 kg N ha^{-1} in most fields and wheat yields in fields with no N applied ranged between 0.2 and 4.8 t ha^{-1} (**Figure 3C**). These N application rates are way lower than the 150 kg N ha^{-1} what is commonly used in on-station trials (Bezabih et al., 2018; Habte et al., 2014) and the N requirements to narrow yield gaps up to the water-limited yield (www.yieldgap.org). The fitted quantile regressions showed a positive relationship between N applied and wheat yields up to ca. 80 kg N ha^{-1}, after which no further responses were observed. When comparing different zones, it is worth noting the greater N applications in Shoa and Gojam than in Arsi and the greater yield responses to N in Shoa, but not Gojam, than in Arsi. P applications rates were smaller or equal to 20 kg P ha^{-1} for most fields and wheat yields in field with no P applied ranged between 0 and ca. 3.5 t ha^{-1} (with only five observations above the later threshold; **Figure 3D**). There was a positive linear relationship between wheat yields and P applied across the highest yielding fields, independently of the percentiles to which the quantile regressions were fitted.

Herbicides were used at rates lower or equal to 2.0 L ha^{-1} in most fields and quantile regressions fitted to the 98th and 90th percentiles showed a positive linear relationship between herbicide use and wheat yields (**Figure 3E**). Hand-weeding is a labour intensive operation as up to 50 person-day/ha were used in some fields and there was relationship between labour used for hand-weeding and wheat yields (**Figure 3F**). It is worth noting that herbicides and hand-weeding were both used in the fields with yields greater than 5 t ha^{-1}. The labor-saving effects of herbicides remain to be tested, which is possible using the available data.

5) Limitations and next steps

As summarized in Appendix A, we encountered a number of data problems in the preliminary analyses of the LSMS which seriously hinder its use for yield gap analysis in this project. Similar problems might also be true for the WAIS data and we will try to assess these through comparisons with secondary sources, literature and expert knowledge. Data quality problems were also encountered in the weather data used for simulating Yp and Yw but we have procedures available, or alternative data sources, to overcome these limitations.

The next step in the analysis of the LSMS and WAIS datasets will focus on identifying the determinants of wheat yields at national and regional levels in Ethiopia. For this purpose, we will use Ordinary Least Squares (OLS) and the potential covariates to include are the variables summarized in Tables 1 and 2 in addition to information on biophysical conditions from GYGA or the datasets themselves. Further, we will apply stochastic frontier analysis to characterize the production frontier, disentangle efficiency and resource yield gaps and identify the drivers of the efficiency yield gap (Silva et al., 2017).

The analyses described above will be done with the WAIS dataset for different zones (regional level) and for the pooled sample (national level). We plan to conduct more detailed analyses in one specific region (e.g., Arsi – Figure 3) so that we can assess the variation observed in the farm data and compare these to experimental data such as the variety trials summarized in

Bezabih et al. (2018). We want to highlight that so far we were not able to compile any experimental data from CIMMYT and we consider this a limitation for this study. If possible, a similar approach at national and regional levels will be followed using the LSMS dataset. This is interesting as it would allow us to assess the robustness, and uncertainties, of the results obtained with the different datasets.

In parallel, actual yields from farmers' fields reported in the surveys will be compared to the simulated yields (and other information) from GYGA. This will be done at a national level, and sub-national levels to the extent allowed by the sampling scheme of the surveys. This last step is necessary to estimate the technology yield gap.

Our ultimate goal is to summarize the findings from the analyses described above in one manuscript on '*Wheat yield gaps in the Ethiopian highlands: Magnitude, drivers and policy implications*'.

6) References

Abro, Z.A.; Jaleta, M.; Qaim, M. (2017). Yield effects of rust-resistant wheat varieties in Ethiopia. *Food security*, 1343 – 1357.

Abro, Z.A.; Jaleta, M.; Teklewold, H. (2018). Does intensive tillage enhance productivity? Panel data evidence from smallholders' agriculture in Ethiopia. *Journal of Agricultural Economics*, 756 – 776.

Bezabih, M.; Adie, A.; Ravi, D.; Prasad, K.V.S.V.; Jones, C.; Abeyo, B.; Tadesse, Z.; Zegeye, H.; Solomon, T.; Blümmel, M. 2018. Variations in food-fodder traits of bread wheat cultivars released for the Ethiopian highlands. *Field Crops Research,* 1 – 7.

Grassini, P.; van Bussel, L.G.J.; van Wart, J.; Wolf, J.; Claessens, L.; Yang, H.; Boogaard, H.; de Groot, H.; van Ittersum, M.K.; Cassman, K.G. 2015. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Research,* 177, 49 – 63.

Habte, D.; Tadesse, K.; Admasu, W.; Desalegn, T.; Mekonen, A. (2014). Agronomic and economic evaluation of the N and P response of bread wheat growing in the moist and humid midhighland vertisols areas of Arsi zone, Ethiopia. African Journal of Agricultural Research, 10, 89 – 99.

Silva, J.V.; Baudron, F.; Reidsma, P.; Giller, K.E. (under review). Is labour a major determinant of yield gaps in sub-Saharan Africa? A study for cereal-based production systems in Southern Ethiopia. *Agricultural Systems*.

Silva, J.V.; Reidsma, P; Laborte, A.G.; van Ittersum, M.K. (2017). Explaining rice yields and yield gaps in Central Luzon, Philippines: An application of stochastic frontier analysis and crop modelling. *European Journal of Agronomy,* 82 Part B, 223 – 241.

Shiferaw, B.; Kassie, M.; Jaleta, M.; Yirga, C. (2014). Adoption of improved wheat varieties and impacts on household food security in Ethiopia. *Food policy*, 44, 272 – 284.

Van Bussel, L.G.J.; Grassini, P.; van Wart, J.; Wolf, J.; Claessens, L.; Yang, H.; Bogaard, H.; de Groot, H.; Saito, K.; Cassman, K.G.; van Ittersum, M.K. 2015. From field to atlas: Upscaling of location-specific yield gap estimates. *Field Crops Research*, 177, 98 – 108.

Van Wart, J.; Grassini, P.; Yang, H.; Claessens, L.; Jarvis, A.; Cassman, K.G. 2015. Creating long-term weather data from thin air for crop simulation modelling. *Agricultural and Forest Meteorology*, 208, 49 – 58.

Disentangling agronomic and economic yield gaps in Ethiopian wheat based systems for better targeting of development interventions (Yield Gap Wheat Ethiopia)

Report #2: Appendix A / LSMS dataset

1) Introduction

In this appendix we discuss the quality of the variables in the Ethiopian LSMS-ISA dataset and justify our choice to use crop cut rather than farmer recall yields. We begin by describing some basic details of the Ethiopian LSMS-ISA dataset. Next, we compare the advantages and disadvantages of crop cut and farmer-recall yield measures. Our experience working with both the Ethiopian and Nigerian LSMS-ISA dataset has highlighted several data quality issues. We describe in general terms the issues we encountered and then use the fertilizer variables and land holdings as specific examples. Our results suggest systematic biases centered on particular enumerators. Finally we assess the crop cut yield data on the same basis and conclude it is more reliable than the available farmer-recall data and can be used in the context of this project to assess wheat yield gaps in Ethiopia.

2) Data

This report uses data from the Ethiopian Living Standards Measurement Survey - Integrated Surveys in Agriculture (LSMS-ISA). The data was collected in three waves corresponding to 2010-11, 2013-14 and 2015-2016. The objective of the LSMS-ISA is "to improve [the] understanding of the inter-relationship between agriculture and poverty reduction, to improve the capacity of national statistics offices to collect and use this data to inform policy, and to foster innovation in the measurement of agricultural data." (Carletto et al., 2015). The survey began with 4000 households who were tracked and re-interviewed in subsequent waves of the survey. The Ethiopian LSMS-ISA is part of the larger LSMS-ISA program to improve the quality of agricultural data through a collaboration between the World Bank and the respective national bureau of statistics in each of the seven LSMS-ISA countries.

Particularly important is the structure in which the data was collected. Enumeration areas were randomly selected for inclusion in the survey and households were then randomly selected from within each selected enumeration area. From the Basic Information Documentation accompanying the LSMS-ISA dataset for Ethiopia we know that enumerators are resident in specific enumeration areas (EAs).

"Resident enumerators were used to administer the household, agriculture (post-planting and post-harvest), and livestock questionnaires in rural and small-town EAs. The CSA assigned one resident enumerator for each of these EAs. The enumerator lived in the EA for the entire survey period from September 2015 to May 2016. Daily laborers/ field guides were also hired locally for a few days to assist the enumerators in parcel/ field measurement and crop cutting activities.

They also helped when child anthropometrics were taken. Temporary mobile enumerator teams were used for the mid- and large-size town EAs." (Basic Information Document wave 3 pp.22)

The quote above is important because patterns of systematic mistakes do exist in the data and they are centered on specific enumeration areas. As enumerators are resident in specific enumeration areas this strongly suggests that certain enumerators make specific mistakes. One option is to find and exclude these enumerators but this comes with several drawbacks. First, there appear to be a considerable number of enumerators who make mistakes in recording quantitative variables. Excluding all enumerators who make mistakes in any variable will take a large toll on the number of remaining observations for analysis. Second, enumerators are located within enumeration areas which are spatially explicit. Excluding particular enumerators means selectively excluding characteristics from that enumeration area such as soil or climate characteristics. Finally, the time required to appropriately clean even a few quantitative variables is prohibitive for this project.

3) Crop cut yields versus farmer recall yields

The LSMS-ISA dataset contains several variables that can be used to construct a wheat yield measure. A crop cutting module recorded both the fresh and dry weight of crop cut on a two metre squared area of the field in wave 1 and four metre squared area in waves 2 and 3. Farmer recall of crop production was recorded for a larger number of plots in the post harvest questionnaire. GPS Field area measurements and farmer self-reported area measurements were recorded for most fields and rope and compass measurements were taken on small plots for which GPS measurements are known to be less accurate (Carletto et al., 2016). We refer to yields calculated by dividing the fresh weight of crop cut by the crop cut area as crop cut yield. We refer to yields calculated by dividing the total farmer reported production by the total field area (measured by GPS, self-reported or rope and compass) as the farmer-recall yield.

$$crop\ cut\ yield = \frac{crop\ cut\ (kg)}{crop\ cut\ area\ (ha)}$$

$$farmer\ recall\ yield = \frac{farmer\ recall\ production(kg)}{field\ area\ (ha)}$$

Both crop cut yields and farmer recall yields have advantages and disadvantages. Here we describe the most important advantages and disadvantages of each measure based on existing literature including an excellent review by Fermont and Benson (2011) as well as particular details which are specific to the Ethiopian LSMS-ISA dataset.

3.1) Advantages of crop cut yields

1. Cut on a marked out area which means the crop cut has virtually no measurement error in the denominator of the crop cut yield (Poate, 1988).
2. Carried out according to a prescribed protocol which is consistent across the entire survey.
3. Carried out by trained enumerators as opposed to household members.
4. Recorded in standard units like kilograms and grams rather than community or region specific local units.

5. The crop cut was weighed fresh, after cutting, and dry roughly two weeks after cutting. Large differences between fresh and dry weight suggests errors making detection of mistakes in the numerator of the crop cut yield simple.
6. The specific date on which crop cutting was carried out is recorded.
7. In theory carried out at the same time as the actual harvest (in practice only if logistically possible).

3.2) Disadvantages of crop cut yields

1. May suffer from several sources of bias (see Fermont and Benson, 2011, for more details):
    - edge effects: Inclusion of plants in the measurement area that actually falls outside it
    - border effects: The tendency for the border of the plot to have a lower chance of inclusion in the measurement area because of the rules governing location.
    - nonrandom location of subplots: tendency of enumerators to consciously or unconsciously avoid low-yielding areas within plots when locating subplots
    - harvest effects: tendency of enumerators to harvest crop cuts more thoroughly than farmers would
    - weighting problems: The use of inappropriate weighing scales
2. Crop cut may not be carried out at exactly the same time as the actual harvest.
3. Typically a smaller sample size due to extra time and cost of crop cutting versus cheaper farmer-recall measures.

3.3) Advantages of farmer recall yields

1. Typically larger sample size because farmer recall is less expensive to carry out compared to crop cutting.

3.4) Disadvantages of farmer recall yields

1. Measured on total field area even when more than one crop is present.
2. Based on farmer recall.
    - Farmers may accidentally misreport yield measures or round production quantities or field areas to the nearest whole unit. For example ten and half bags of harvested wheat may be rounded down to ten bags of wheat.
    - Farmers may deliberately under or over report certain quantities. For example, farmers who pay for farm services or taxes by land holdings may have a tendency to under-report area sizes to avoid costs.
3. Reported in non-standard units which vary by community and region and may not accurately reflect true production quantities.

Assessed on the basis of the advantages and disadvantages of each measure we conclude that crop-cut is preferred. Most of the disadvantages of crop cutting are either minor or mitigated by the specific crop-cutting protocol in the Ethiopian LSMS-ISA. For example, the non-random location of subplots within an area of crop is largely mitigated in the Ethiopian LSMS-ISA because the crop cutting protocol has a procedure to randomly select a subplot based on the dimensions of the full field.

4) The quality of agricultural data

We now turn to a more specific assessment of the quality of the actual data recorded in the Ethiopian LSMS-ISA. Several recent studies have examined the quality of aspects of data gathered as part of the wider LSMS-ISA project or in household survey data more generally. For example, Beegle et al. (2012) study the impact of recall bias on survey data, Carletto et al. (2013) and Carletto et al. (2015) study the effect of GPS measured land holdings in comparison to self-reported field holdings and Gourlay et al. (2017) and Desiere and Jolliffe (2018) investigate the relationship between crop-cut based production estimates and self-reported crop production data in Uganda and Ethiopia respectively. In the case of Desiere and Jolliffe (2018), the same LSMS-ISA dataset used in this report was also used in that paper.

More recently, a number of studies have assessed the credibility and reliability of survey data by means of statistical analysis using certain plausibility bounds (Faval et al., 2018) or by trying to model the systematic error explicitly (Abay et al., 2018; Faval et al., 2018; Tasciotti and Wagner, 2017).

One criticism of the existing literature is that it tends to approach the issue of data quality at quite a high level. What is meant by this, is that existing literature either investigates the techniques used to gather the data, broadly compares existing data with plausibility bounds, or compares more than one dataset for consistency. These techniques offer little practical explanations for what causes the low quality of the data which is often observed by users of the data. We offer a different perspective by focusing on the consistency of the individual data points within one dataset.

For example, the price of fertilizer in Ethiopia is tightly controlled by the Ethiopian government and well reported by several sources (Rashid et al., 2013). We would expect to see similar fertilizer prices reported by farmers across the country. Because transport costs can lead to somewhat different costs in more or less remote regions, comparisons within enumeration areas (EAs) or between neighbouring EAs would give an indication of whether the data has some measure of reliability. We certainly would not expect the cost of inputs to differ between farmers by several orders of magnitude anywhere in the country.

The LSMS-ISA survey for Ethiopia is a panel survey tracking the same households and farms across time. Presumably there should be some consistency in the size of the land holdings of a household when compared over time. Fertilizer quantities and costs are recorded in the LSMS-ISA dataset. We would expect farmers to pay similar unit prices for the same fertilizer products, and that these prices are broadly in line with fertilizer prices in Ethiopian in general. A key finding is that suspicious patterns do exist in the data and these are idiosyncratic at the enumeration area level. We give examples of these patterns in terms of the area and fertilizer variables.

It is not possible in this Appendix to point out all the deficiencies in the data recorded in the LSMS-ISA survey for Ethiopia. Instead we focus on these two examples as indications of where consistency in the data is lacking.

4.1) Fertilizer variables

Fertilizer quantities (kg) and costs (Birr) are recorded in the data at the field level. Figure A1 plots the cost of urea fertilizer in the second and third waves of the data as a function of the quantity of urea purchased for several EAs. Each EA is chosen for illustrative purposes and similar figures for all EAs are available on request. The dotted black line connecting the quantity and cost is a price line of 12 Birr/kg and it is added for orientation. This line is in accordance with external sources of information on fertilizer prices in Ethiopia.

Because EAs are roughly equivalent in size to villages we would expect that each observation would fall close to the price line and be consistent across EA. This is true for the EAs in the top left and bottom right panes of Figure A1. In each of these EAs farmers report urea costs and quantities which are in line with external information on urea prices and have an application rate which is between 50 and 200 kg/ha. It is difficult to distinguish small and large application rates using the legend in Figure A1 precisely because the misreported values are 100 or 1000 times too large so that capturing all values on one legend is impossible. The fact that such large values exist in the first place is the main purpose of Figure A1.

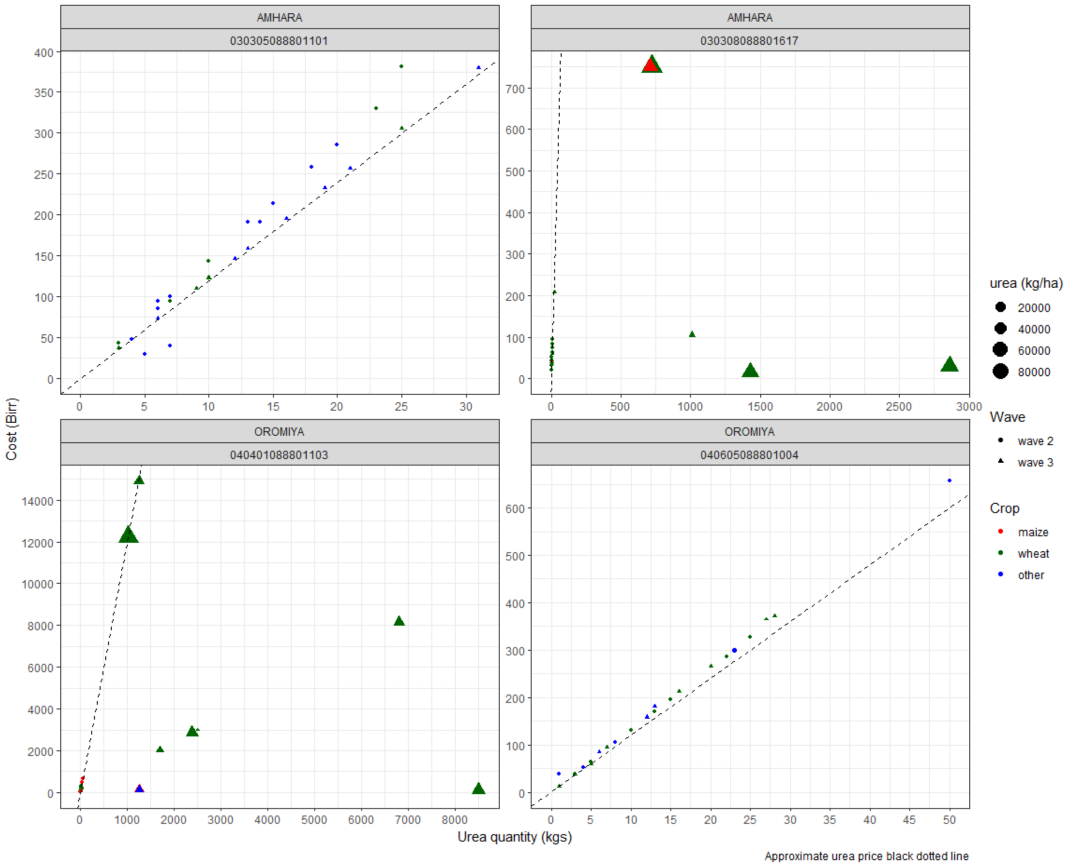

Figure A1. Urea quantity, cost and application rate for selected enumeration areas. These enumeration areas are shown for illustrative purposes.

In comparison, the top right and bottom left panes of Figure A1, indicate that some farmers were ostensibly able to acquire several thousand kilograms of urea at a fraction of the price of neighbouring farmers. Their application rate regularly exceeded 20 tonnes per hectare. In reality, most of these values are simply reported ten or 100 times too large, a common mistake in the Ethiopian LSMS-ISA survey across a large number of variables. Unfortunately, this makes input quantities and reported production unreliable. It is difficult to give a figure of exactly how many plots suffer from this and similar problems but it is substantial.

4.2) Total farm areas

To test the consistency of total farm areas across time we first excluded any farm that was not interviewed in all three waves. Next, we excluded any farm which reported a new parcel relative to the previous survey in any wave. We also removed any farm which reported no longer owning a parcel, relative to the previous survey wave, or had a plot which was not measured by either GPS or rope and compass in any year. The remaining 954 farms should have stable and consistent total farm areas across time. We focus on the 665 farms located in the four largest regions; Amhara, Oromiya, SNNP and Tigray. These are the most relevant regions for wheat production.

Figure A2 plots the wave 2 total farm areas against the wave 1 and wave 3 total farm areas in hectares. If the total farm area is consistent across time we would expect all points to fall on the 45 degree black line. Although we see a rough upward trend, in general the total farm areas in each region do not fall on the 45 degree line. This suggests that either there is a problem with the data or household land holdings are extremely fluid over time.

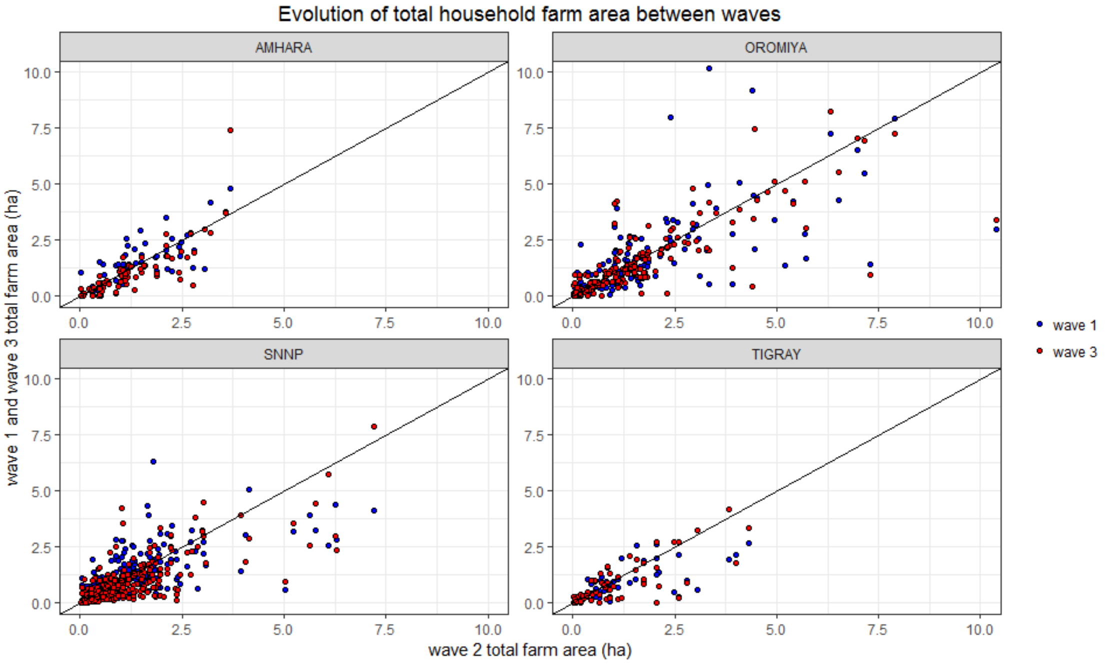

Figure A2. Comparison of total farm land holdings across time for selected regions.

The percentage change in wave 3 land holdings relative to wave 2 can be defined by

$$\%\Delta land_{w3,w2} = \frac{land_{w3} - land_{w2}}{land_{w2}}$$

However, because either the wave 2 or the wave 3 land holdings could be incorrect, we also calculate the percentage change in wave 2 land holdings relative to wave 3 land holdings

$$\%\Delta land_{w2,w3} = \frac{land_{w2} - land_{w3}}{land_{w3}}$$

We plot both $\%\Delta land_{w3,w2}$ and $\%\Delta land_{w2,w3}$ in Figure A3. Relative to wave 2, 31% of wave 3 total farm areas have either increased or decreased by at least 50% while relative to wave 3, 38% of wave 2 total farm areas have increased of decreased by at least 50%.

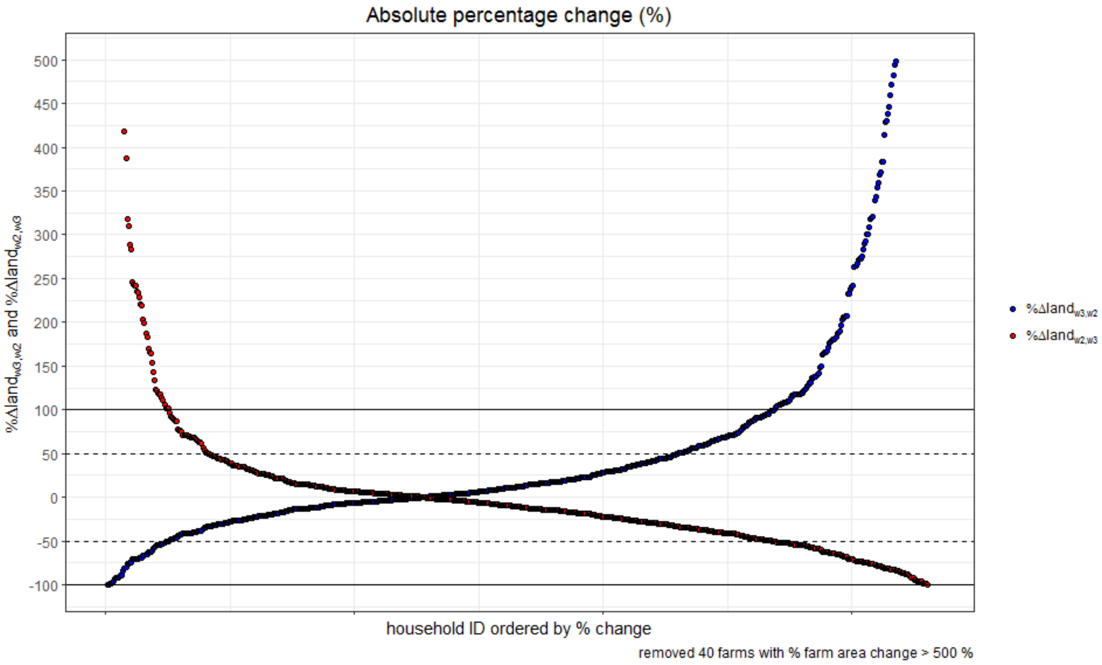

Figure A3. Percent change in total farm land holdings between waves 2 and 3. Only farms above 500% change are excluded, as a result it may appear that some farms only appear once - their corresponding value is outside the plot.

It is not impossible that total farm area changes for reasons that escape the survey questions. For example, if land shared by separate but related families is pooled in some seasons, it may appear that household land holdings change between waves. However, a closer look at the evolution of land holdings across time suggests this is not the case. Figure A4 tracks the evolution of land holdings across four enumeration areas. In the top two panes of Figure A4 we see relatively consistent total farm holdings across each wave of the survey; just as we would expect. In the bottom two panes of Figure A4 we see a different pattern. Two households in enumeration area 041712088800807 see a substantial increase in their total farm land holdings in wave 2. This increase then disappears by wave 3. Similarly, the bottom pane of Figure A4 suggests that farmers in ea 070112088802605 maintain relatively consistent total land holdings between waves 1 and 2 before most farmers loose significant amounts of land by wave 3. In the absence of a good explanation for why these changes in total land holdings occur, we assume they are sue to some kind of measurement error which is likely to affect area denominated measures of wheat productivity and input use. Similar figures are available for all enumeration areas on request.

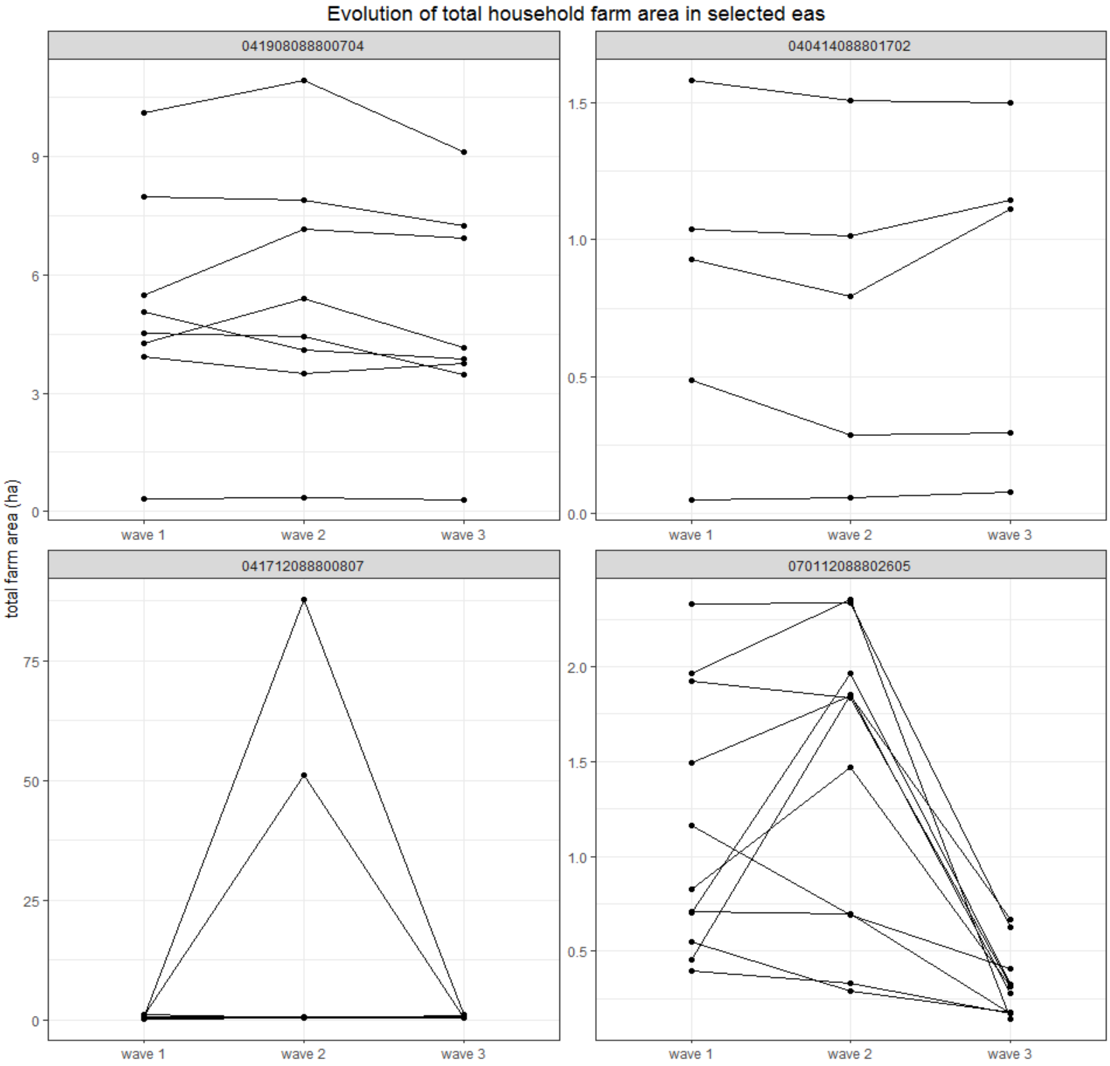

Figure A4. Evolution in total land holdings between waves 1 and 3 for selected enumeration areas.

5) Conclusion

The purpose of this Appendix is to justify why we choose crop cut as a measure of wheat yield in Ethiopia as opposed to wheat yields based on farmer recall of crop production and field area size. This choice can be justified by discussing the merits and disadvantages of both measures. We conclude that the major disadvantages of crop cut as a measure of yield (edge effects, border effects and the non-random location of subplots) are unlikely to pose a problem in the LSMS-ISA dataset because of the protocol followed by enumerators in gathering the crop cut data.

On the other hand, we have described several serious deficiencies in the fertilizer and field area measurements which lack consistency when compared to external information, other measurements within the same survey or common sense. We expect most quantitative information drawn from this survey to have similar issues which we have also seen in several other surveys collected in, for example, Nigeria and Ghana. As a result, we focus our attention on the crop cut yields, binary and categorical information on inputs and crop management from within the survey, and geographical information on growing conditions including soil and climate characteristics.

6) References

Abay, K.A.; Gashaw, T.A.; Barrett, C.B.; Tanguy, B. 2018. Correlated Non-Classical Measurement Errors, ‘Second Best’ Policy Inference and the Inverse Size-Productivity Relationship in Agriculture. Vol. 1710. International Food Policy Research Institute.

Beegle, K.; Carletto, C.; Himelein, K. 2012. Reliability of recall in agricultural data. *Journal of Development Economics*.

Carletto, C.; Jolliffe, D.; Banerjee, R. 2015. From Tragedy to Renaissance: Improving agricultural data for better policies. *The Journal of Development Studies,* 51 133–48.

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, 254–61.

Desiere, S.; Jolliffe, D. 2018. Land productivity and plot size: Is measurement error driving the inverse relationship? *Journal of Development Economics*.

Fermont, A.; Benson, T. 2011.Estimating yield of food crops grown by smallholder farmers: A review in the Uganda context.

Carletto, G.; Gourlay, S.; Murray, S.; Zezza, A. 2016. Land area measurement in household surveys: A guidebook. Washington DC: World Bank.

Poate, D. 1988. A review of methods for measuring crop production from smallholder producers. *Experimental Agriculture*.

Rashid, S.; Dorosh, P.A.; Malek, M.; Lemma, S. 2013. Modern input promotion in sub-Saharan Africa: Insight from Asian green revolution. *Agricultural Economics*, 44, 705 – 721.

Tasciotti, L.; Wagner, N. 2017. How Much Should We Trust Micro-data? A Comparison of the socio-demographic Profile of Malawian households using census, LSMS and DHS data. *The European Journal of Development Research,* 1–25.

Disentangling agronomic and economic yield gaps in Ethiopian wheat based systems for better targeting of development interventions (Yield Gap Wheat Ethiopia)

Report #2: Appendix B / WAIS dataset

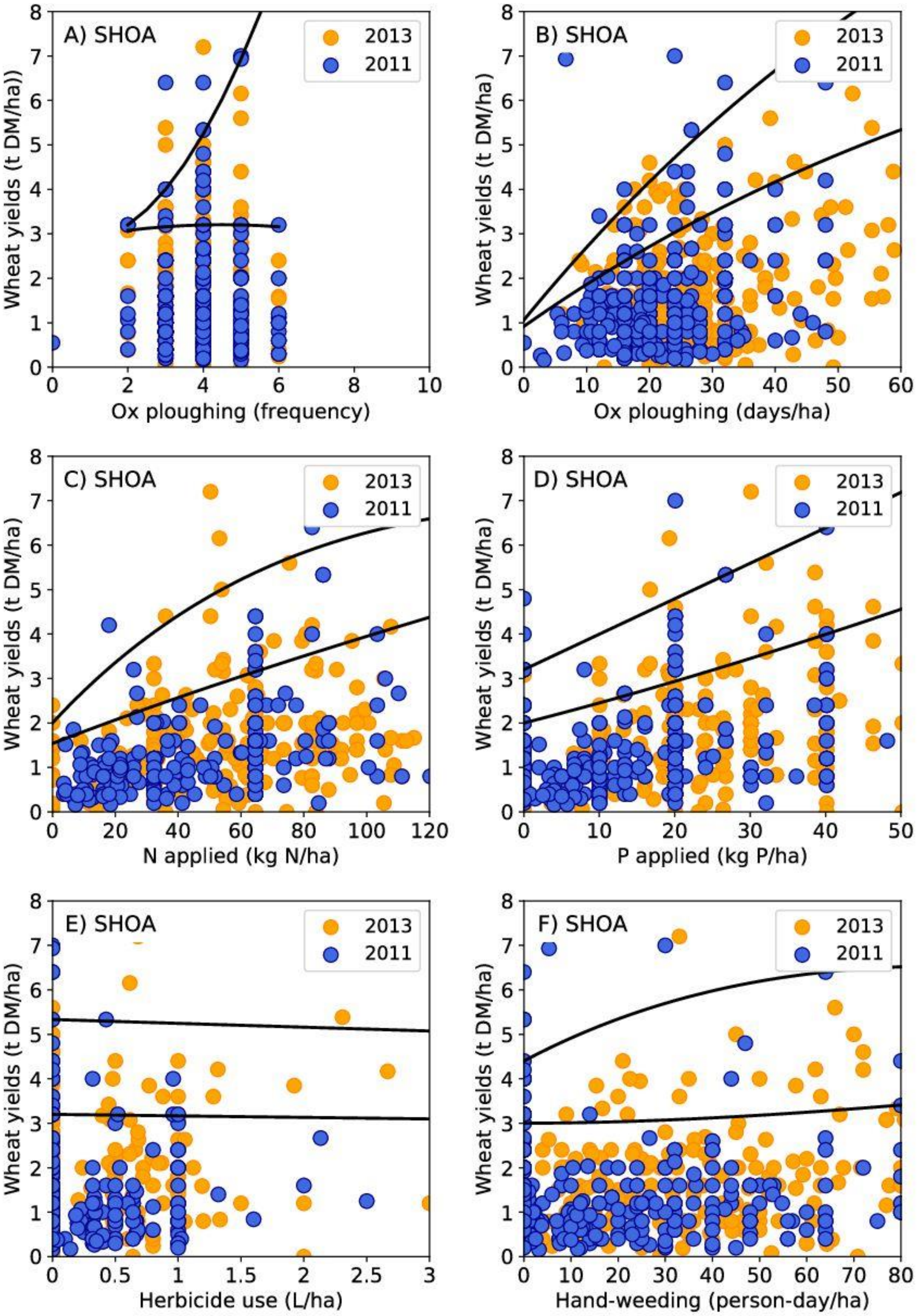

Figure B1. Wheat yield responses to A) ox ploughing frequency, B) ox ploughing days, C) N applied, D) P applied, E) herbicide use and F) hand-weeding in East and North Shoa zones, Ethiopia. Data are presented for the Meher season only. Solid lines show fitted quantile regressions (y = a + b × 0.99^x + c × x) to the 98th (top) and 90th (bottom) percentile. The 98th percentile capture the influence of extreme observations while the 90th percentile refers to the lower threshold for highest farmers' yields commonly used.

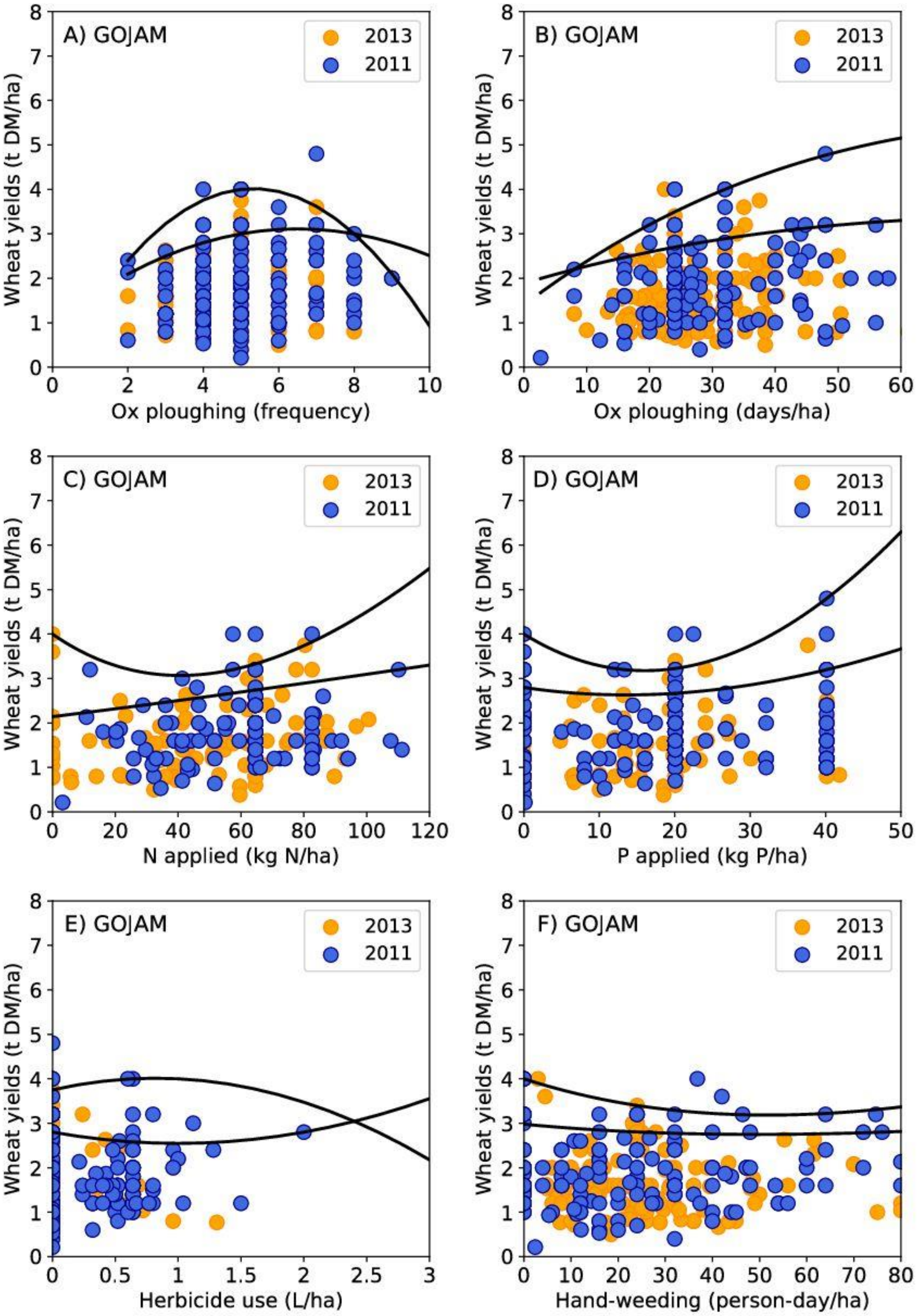

Figure B2. Wheat yield responses to A) ox ploughing frequency, B) ox ploughing days, C) N applied, D) P applied, E) herbicide use and F) hand-weeding in East Gojam zone, Ethiopia. Data are presented for the Meher season only. Solid lines show fitted quantile regressions ($y = a + b \times 0.99^x + c \times x$) to the 98th (top) and 90th (bottom) percentile. The 98th percentile capture the influence of extreme observations while the 90th percentile refers to the lower threshold for highest farmers' yields commonly used.