Perceived, measured, and estimated soil fertility in east Africa: Implications for farmers and researchers

Julia Berazneva, Department of Economics, Middlebury College (jberazneva@middlebury.edu)

Linden McBride, Dyson School of Applied Economics and Management, Cornell University (lem247@cornell.edu)

Megan Sheahan, Precision Agriculture for Development (msheahan@precisionag.org)

David Güereña, Once Acre Fund (david.guerena@oneacrefund.org)

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2.

Perceived, measured, and estimated soil fertility in east Africa: Implications for farmers and researchers

Julia Berazneva[1], Linden McBride[2], Megan Sheahan[3], and David Güereña[4]

[1] *Department of Economics, Middlebury College (jberazneva@middlebury.edu)*
[2] *Dyson School of Applied Economics and Management, Cornell University (lem247@cornell.edu)*
[3] *Dyson School of Applied Economics and Management, Cornell University (mbs282@cornell.edu)*
[4] *Once Acre Fund (david.guerena@oneacrefund.org)*

March 2016

Abstract

Bringing together emerging lessons from biophysical and social sciences as well as newly available data, we take stock of what can be learned about the relationship among perceived soil fertility, measured soil fertility, and farmer management practices in east Africa. We identify the correlates of Kenyan and Tanzanian maize farmers' reported perceptions of soil fertility and assess the extent to which these subjective assessments reflect measured soil chemistry. Our results offer evidence that farmers base their perceptions of soil quality and soil type on crop yields. We also find that, in Kenya, farmers' reported soil type is a reasonable predictor of several objective soil fertility indicators while farmer-reported soil quality is not. In addition, in exploring the extent to which publicly available soil data are adequate to capture local soil chemistry realities, we find that there is still immense value to the time-consuming collection of soil samples where highly accurate soil measures are important to research objectives. However, in the estimation of agricultural production or profit functions, where the focus is on averages and where there is low variability in the soil properties, there may be limited value to including any soil information in the analysis.

Keywords: natural resource management, soil fertility, agricultural productivity, farmers' perceptions, Kenya, Tanzania.

JEL codes: O13, Q12, Q24, Q56.

Acknowledgements

For valuable input and comments on an earlier draft, we thank Christopher Barrett, Anil Bhargava, David R. Lee, Bart Minten, Cheryl Palm, Emilia Tjernström, and participants at the Second Global Food Security Conference in Ithaca, NY and the Structural Transformation of African Agriculture and Rural Spaces (STAARS) Conference in Addis Ababa, Ethiopia. The data collection in Kenya was supported by the Atkinson Center for a Sustainable Future and the World Agroforestry Center; funding was provided by *Fondation des Fondateurs*, Cornell University, and an AAEA Tweeten Scholarship. The authors thank Siobhan Murray, Amparo Palacios-Lopez, and others on the LSMS-ISA team at the World Bank for their insights on the Tanzania data.

1 Introduction

While many socio-economic factors contribute to poor crop yields across Sub-Saharan Africa (SSA), a major biophysical contributor is the depletion of soil fertility (Sanchez 2002; Sanchez and Swaminathan 2005; Vanlauwe et al. 2015; Tully et al. 2015). Across different agro-ecological zones in SSA, soils poor in nutrients and soil organic matter not only partially account for low yields but also limit the effectiveness of other inputs such as fertilizer and labor, and reduce farm household resilience to external stressors and shocks (e.g., pests, crop diseases, climate change). Moreover, the direct links between soil fertility, agricultural productivity, food insecurity, and rural poverty can be self-reinforcing. Whether due to an initial poor soil endowment or resource constraints leading to low input use (fertilizers and/or organic soil amendments), the broad pattern across much of SSA is soil degradation over time (Tittonell et al. 2005a; Guerena et al. 2016). As a result, some farmers find themselves trapped in a low productivity equilibrium (Shepherd and Soule 1998; Antle et al. 2006; Stephens et al. 2012; Barrett and Bevis 2015). Despite the importance of soil fertility in the context of agricultural development, major barriers remain in our understanding of how farmers form perceptions about their soil fertility, and how soil fertility (perceived, directly measured, and estimated) is related to farmers' behaviors in terms of input use, cropping strategies, and other management practices.

A paucity of research directly examines the relationship between soil fertility and the existing farm management practices, especially in SSA. Agronomic studies that have precise measures of soil fertility and yields often ignore farmers' behavioral responses (see, for example, Vanlauwe et al. (2011)) while economic studies fail to account for soil fertility in estimation of agricultural profits and farmer welfare, at best including indicator proxy variables for soil fertility (e.g., Duflo et al. (2008); Sheahan et al. (2013)). Only several studies with access to precise measures of soil fertility analyze farmers' knowledge of land quality and within-farm variability in resource allocation and yields (e.g., Tittonell et al. (2005b)). Therefore, in this paper, we attempt to bring together emerging lessons from the biophysical and social sciences as well as newly available data to take stock of what we can learn about the relationships among perceived (subjective), measured, and estimated soil fertility and farmers' management practices.

Several other studies have examined these relationships, with mixed results. Cross-sectional data from the World Bank's Living Standards Measurement Study-Integrated Survey in Agriculture (LSMS-ISA) surveys across six different countries, for example, suggest that farmers in SSA do not significantly vary input application rates according to perceived soil fertility (Sheahan and Barrett 2014). At the same time, there is evidence from Kenya that farmers apply fewer external inputs on soils with objectively verified low soil carbon content and fertility (Marenya and Barrett 2009a), and adjust planting timing and weeding intensity on plots with different land quality (Tittonell et al. 2005b). To better understand these empirical observations, we identify the correlates of farmers' perceptions of soil fertility and assess whether currently available soil data are adequate to capture farmers' perceptions and knowledge.

We also explore the extent to which publicly available soil data, estimated via sophisticated interpolation methods from point observations across the African continent, are adequate to capture local soil chemistry realities at the household, village, and sample levels. Such data sets are an incredible resource and their availability may obviate the need for detailed on the ground soil data collection, saving researchers, agricultural organizations, and governments both time and money. This exercise allows us to make recommendations to the broader research community about the relative trade-offs inherent in relying on one soil metric over another. Finally, we assess the value of soil information from a research standpoint by interchanging various soil metrics in a production function approach to the analysis of yields.

In particular, we address the following four research questions in the course of our analysis:

1. What can we learn from household survey data about how farmers in east Africa form perceptions about their soil fertility? Do agricultural inputs and/or outputs vary with perceived soil quality and soil type?

2. How well do farmers' subjective perceptions of soil fertility correspond to objective laboratory measurements of soil chemistry? Are there any systematic factors affecting farmers' soil fertility perceptions?

3. Can the new high-resolution and publicly available estimated soil fertility data sets provide sufficient information to obviate the expensive and time-consuming collection of detailed plot-level data?

4. What is the value of soil information? Are there significant costs to farmers' and researchers' (mis)perceptions about local soil fertility?

To answer these questions, we rely on three data sets that correspond with a small number of maize farming households in western Kenya and two data sets that correspond with a nationally representative sample of maize farmers in Tanzania. In both study regions, farmers' perceptions of soil fertility and their agricultural practices are drawn from household survey responses. Global positioning system (GPS) coordinates allow us to match all of these households with publicly available geo-referenced soil data at 250-meter spatial resolution from the Africa Soil Information Service (AfSIS) (Hengl et al. 2015). In western Kenya, additional laboratory measures of plot-level soil fertility are obtained from soil analysis based on the resource- and time-intensive collection of soil samples (Berazneva et al. 2016). Apart from geographic differences, both the Kenya and Tanzania data sets also offer different contexts in terms of data collection efforts: the Kenya data are from a small-scale detailed survey, while the Tanzania data are from a nationally representative large-scale project. Combining the two geographic locations allows us to compare across the contexts, provide external validity to our findings, and offer recommendations to researchers on soil data collection and use.

Our results offer some evidence that farmers base their perceptions of soil quality and soil type on crop yields. We also find that, in Kenya, farmers' reported soil type (soil texture) is a reasonable predictor of several objective soil fertility indicators drawn from laboratory soil analysis while farmer-reported soil quality is not. In addition, we find that the differences between the two objective soil data sets that we compare in Kenya—plot-level measured soil analysis data and estimated AfSIS soil data—are considerable, indicating that there is still immense value to the time-consuming collection of soil samples where highly accurate and local soil measures are important to research or extension objectives, despite the growing availability of high-resolution geo-referenced soil data. However, in the estimation of agricultural production or profit functions, where the focus is on averages and where there is low variability in the soil properties, there may be limited value to including any soil information in the analysis.

Our paper proceeds as follows. In the next section, we briefly discuss the context from which our research questions arose. We then discuss our data sources and the methods, mostly descriptive, we use. The following section offers results for each of the four questions under investigation. The last section concludes, taking stock of what we have learned about the relationship between perceived, measured, and estimated soil fertility and farmer management practices, and offering additional research questions worth pursuing, both for better comprehension of farmer behavior and for the collection of better data.

2 Background

The international development community has recently begun to turn its attention towards the role of soils in agricultural and human development; in fact, the Food and Agriculture Organization of the United Nations declared 2015 the International Year of Soils. Aware that soils are important, development and agricultural economists are increasingly including soil fertility data in their analyses. When it comes to using soil data, economists generally fall into three camps. The first camp takes farmers' knowledge of soil fertility as a sufficient measure of or sufficient proxy for soil fertility without any verification exercise or follow-up discussion about how farmers make these determinations (see, for example, Sherlund et al. (2002)). These measures are often used as control or proxy variables. The second camp assumes that farmers are too information-constrained to accurately report soil quality measures and therefore relies on highly aggregated or estimated measures of soil quality or soil type, derived from external mapping exercises and often matched using administrative boundaries (e.g., Sheahan et al. (2013)). The third camp makes the same assumptions as the second but collects and analyzes soil samples from the actual plots or farms under study in lieu of relying on highly aggregated or predicted external data sets (e.g., Marenya and Barrett (2009a)).

Each camp makes reasonable assumptions under the reality of data constraints, but little research attempts to empirically understand the uniqueness of the information embodied in each of these types of soil data. This information is valuable when choosing the most accurate soil fertility metrics for analysis and in understanding the reasons why other metrics may be insufficient. This is particularly true of farmer-reported values, which are collected in most agricultural surveys implemented today.

While it is reasonable to expect that farmers in SSA are constrained in their ability to know the precise nutrient content of their soils, farmers do form assessments of their soil fertility and productivity (Niemeijer and Mazzucato 2003). To our knowledge, only a few studies in economics have sought to aid our understanding of this process absent measurement.[1] Marenya et al. (2008), for example, study farmers' perceptions of soil fertility and the impacts of fertility on yields in western Kenya. Using objective measures of soil fertility, the authors find evidence of widespread farmer misperception of soil fertility and these misperceptions cannot be easily explained by observed plot or farmer characteristics such as plot size or farmer's gender or age. The Kenyan farmers in the study, similar to the farmers of the south-central highlands of Ethiopia (Karltun et al. 2013), use crop yields as the key soil fertility indicator. Yet if yield changes lag behind the changes in soil fertility, farmers may be unable to identify important dynamic patterns in soil fertility and may be slow to update their assessments. This delayed response can result in significant deterioration in soil fertility or render soils unresponsive, making regeneration efforts expensive. Once soil has degraded below a productivity threshold, soil restoration can become prohibitively costly and therefore "economically irreversible" (Antle et al. 2006).

Moreover, resource allocation and crop management can differ according to perceived soil fertility. Tittonell et al. (2005b), for example, find differences in the timing and intensity of crop management according to farmers' perceptions of the quality of their land in Kenya. More fertile plots are planted earlier, with more spacing between plantings, and are weeded more often. These practices unsurprisingly lead to greater yields. Therefore, subjective soil fertility perceptions matter. However, beyond these few papers, the formation of farmers' soil fertility assessments as well as the interactions between farmers' assessments and land management practices have not been explored.

The formation of farmers' perceptions about their soil fertility and the farming practices that flow from these perceptions are important to understanding the critical linkages between resource endowments, crop and land management, and agricultural productivity. These linkages, in turn, may have major policy and programmatic implications. From a research perspective, understanding the formation of farmers' soil assessments is a first step towards evaluating the research value of these subjective measures.

[1] A review of rural development literature, as well as studies in ethnopedology that focus on how farmers understand their soils based on collective experiences, can be found in Marenya et al. (2008).

If objective measures of soil fertility are deemed preferable over subjective measures, then the next logical question is whether researchers should forsake free and publicly available data sets in pursuit of their own expensive and time-consuming collection of soil chemistry data; i,e, which of the soil-data-using-economist-camps is preferred? Massive amounts of resources have been funneled into the creation of these publicly available data sets with high resolution and either continental or global coverage, including but not limited to AfSIS[2] and the FAO's Harmonized World Soil Database.[3] In fact, the publicly available household survey data collected by national statistics agencies throughout SSA and overseen by the LSMS-ISA initiative include files with soil data from the FAO. Researchers wanting an even fuller complement of soil variables can easily match the household survey data with the AfSIS database using provided enumeration level coordinates. But, in the end, these soil data sets are the result of interpolation and are only as good as the data fed into the algorithm and the underlying model. Moreover, interpolation itself implies that the areas between sampling points are estimated, which, depending on the spatial resolution of the data, may have large associated error. Without a critical assessment of how well these data represent local soil chemistry realities, as derived from plot-level soil analysis, researchers cannot make good decisions about which data may be most appropriate for their work. With very few exceptions (e.g., Bui (2010)) comparative analyses of a publicly available spatial soil databases with plot level soils data are not available, and we have found no studies that assess the performance of AfSIS at the local level.

With renewed international recognition of the important role soils play in agricultural production, welfare dynamics, and carbon sequestration (Lal 2012; Barrett and Bevis 2015; Lehmann and Kleber 2015) as well as with major resources being devoted to the collection of a variety of subjective and objective, measured and estimated indicators of soil fertility, it is imperative to assess what these data can and cannot tell us. This paper helps to sort through the implications by bringing together and comparing some of these data sources.

[2] http://www.isric.org/data/afsoilgrids250m

[3] http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/

3 Data and methods

Since crop choice may be both a function of and response to a farmer's perceived soil fertility, we limit our analysis to maize, the main and most important cereal in east Africa. Our data come from Tanzania and western Kenya and are described in the subsections that follow. After providing details on the data, we describe the analytical methods used to answer our four research questions.

Farmer-reported soil fertility measures, yields, inputs

We use a nationally representative sample of households from the 2010-2011 wave of the publicly available Tanzania National Panel Survey, data collected as part of the World Bank's LSMS-ISA project. From the full sample, we restrict our analysis to the sub-sample of 2,360 plots containing maize in the main growing season across 1,566 households, with plot-level data on agricultural inputs and maize yield. A typical LSMS-ISA questionnaire asks respondents to specify soil quality and type for each plot under cultivation without prompting or guidance so that farmers' responses should be purely based on their perceptions. The responses are then grouped into pre-coded categories. For the Tanzania LSMS-ISA, the pre-coded categories for soil quality are good, average,[4] or bad; for soil type or texture they are sandy, loam, clay, or other. The presence of sampling weights allows us to apply household-level population weights in the statistical analysis that follows.

The standard modules of the LSMS-ISA questionnaire were adopted for a household survey effort of over 300 households collected in 2011-2012 in fifteen villages in the Nyando and Yala river basins of rural western Kenya (Berazneva et al. 2016). We use data for all maize-growing households for which soil analysis is available, for a sample size of 509 maize plots cultivated by 307 households. Identical to the LSMS-ISA survey, respondents classify their soil quality and type based on their knowledge, as well as report agricultural input and maize production levels. The near-identical questions and classifications between the LSMS-ISA survey implemented in Tanzania and that implemented in Kenya allow us to easily compare across the two regions.[5]

Agricultural input and output variables are drawn from farmer recall related to the last main

[4] We use good, average, and bad soil categorizations to mirror the questions in the household surveys. Average should be understood as intermediate (not an arithmetic mean).

[5] The soil type question was identical across the two data sources. The soil quality question offered several additional pre-coded options (poor, very poor, and not productive at all) in Kenya that were later grouped into the bad category to correspond to the Tanzania data.

season. Where applicable, we standardize input and output values by plot size. For Kenya, plot area is measured with hand-held GPS units. For Tanzania, GPS-measured plot areas are only available for a sub-set of all plots, so we rely on imputed plot sizes as described in Palacios-Lopez and Djima (2014). We also create a range of plot- and household-level characteristics from the survey data, relying mainly on those variables observed consistently across the two countries.

Researcher-collected plot-level soil samples

In western Kenya, soil samples were collected from the largest maize plot of each farm household at the end of the long rains season of 2011. Topsoil (0-20 cm) was randomly sampled from four points across the plot, mixed together (homogenized), and analyzed at the World Agroforestry Center's Soil-Plant Spectral Diagnostics Laboratory in Nairobi, Kenya. The samples were analyzed using mid-infrared spectroscopy (MIR), a rapid nondestructive technique for examining the chemical composition of materials (Shepherd and Walsh 2002; Cozzolino and Moron 2003; Shepherd and Walsh 2007). The MIR analysis provided information on several key soil characteristics: soil carbon measured as percentage of total soil carbon by mass (% by weight or % w/w),[6] nitrogen content measured as percentage of total nitrogen in the soil by mass (% by weight or % w/w), soil pH (measured on 1 to 7 scale), and cation exchange capacity (CEC) measured in milliequivalent of hydrogen per 100 grams of dry soil (meq/100g). Soil carbon and nitrogen content have been used as proxies for soil fertility in the past (see, for example, Marenya and Barrett (2009b)). These two measures are highly collinear and correspond to soil organic matter content that can be transient and influenced by farm management practices. Soil pH and CEC, on the other hand, relate more strongly to soil texture and mineralogy, and therefore are more stable indicators of soil fertility (Sparks 1996).

In order to classify soils as "fertile," we use thresholds and recommendations for soils in western Kenya from the Kenya Agricultural Research Institute (Mukhwana and Odera 2009) and from the Cornell Soil Health Test (Moebius-Clune et al. 2011). Fertile soil is defined as soil with organic carbon content greater than or equal to 2% w/w, total nitrogen content greater than or equal to 0.2% w/w, and pH greater than or equal to 5.2. The resulting soil data offer on-the-ground insight

[6] The soils in the research site in Kenya are acidic and do not contain carbonates so that total stocks of soil carbon are equivalent to total organic carbon content.

into the plot-level soil fertility of smallholder farmers in rural western Kenya. In the discussion below, we refer to these laboratory measurements as "measured soil data" or "soil analysis data."

Geo-referenced and estimated soil quality measures

We also match the household survey data with publicly available data from the Africa Soil Information System (AfSIS). AfSIS, a collaborative soil ecosystem services project, provides data on soil characteristics at 250-meter spatial resolution (Vagen et al. 2010). The data were created by interpolating soil characteristics from more than 28,000 sampling locations using techniques detailed in Hengl et al. (2015).

The AfSIS data were downloaded from the Africa Soil Profiles Database Version 1.2, where .tifs of a variety of soil characteristics are available at 0–5cm, 5–15cm, 15–30cm, and etc. depths. Because we want the AfSIS data to be comparable to the laboratory measured soil data in Kenya, we selected the data representing the 0-20cm depth where available (total soil nitrogen). Where the 0–20cm-depth level was not available (soil organic carbon, pH, and CEC), we selected data representing the 0–5cm and 5–15cm depths and averaged them together.

We paired the AfSIS data with the Kenyan and Tanzanian households by extracting the geo-references available in the household surveys. In Kenya, these points pertain to plots; in Tanzania, these points pertain to the average of the enumeration area (EA), as per World Bank LSMS-ISA restrictions.[7] While the AfSIS data repository provides information about a large number of soil indicators, we extract only the soil characteristics that best match those same values available in the soil analysis data in order to make valid comparisons: soil organic carbon, total soil nitrogen, pH, and CEC.[8] While organic carbon and total nitrogen content are susceptible to change over time, soil pH and CEC are more stable and therefore potentially more appropriate measures of soil fertility to obtain through satellite data.

[7] In order to pair the AfSIS data with the Kenya plot level geo-references, we extracted the values for each soil characteristic as observed (i.e., we extracted the value for the 250m cell the geo-referenced point fell in). So as to pair the AfSIS data with the Tanzania enumeration area geo-references, we extracted the values for each soil characteristic as interpolated (i.e., we extracted a value produced via interpolation from the values of the four nearest raster cells in the AfSIS data). We took these two different approaches—strict extraction versus interpolation—for the two countries due to the nature of the geo-references available to us in the household survey data for each country. However, it should be noted that there was little substantive difference between the observed and interpolated points in either country. For eight EAs in Tanzania, the included geo-reference details landed in bodies of water so that we were unable to match these with AfSIS data. In these cases, we replaced with median values across EAs within a ward.

[8] Carbon: $A/10$, nitrogen: $A/10$, pH: $A/10$, CEC: no conversion necessary as the AfSIS data are already in the same units as the soil analysis data. A indicates AfSIS data.

Analytical methods

We combine the three aforementioned data sets to address our research questions. Graphically, Figure 1 displays the sample of Kenyan households with the soil analysis data (in circles) overlaid on the AfSIS soil pH data. From this figure we can see the relative resolution of the two data sets. In the left panel we observe all fifteen-study villages as well as the general variation in soil pH across western Kenya. Zooming in on one of the villages (in the Lower Nyando region) in the right panel, we see that the variation in soil pH both decreases and becomes more pixelated as we approach the 250-meter resolution level.

Our statistical analysis relies mainly on difference-in-means tests. To determine whether the means of perceived soil quality and soil type differ significantly across agricultural inputs, maize yield, and the measured (soil analysis) and estimated (AfSIS) indicators of soil fertility, we use the Tukey-Kramer test, which allows for multiple pairwise comparisons while accounting for the family-wise error rate. To explore the heterogeneity in farmers' perceptions we also estimate an ordered probit model with a similar set of variables included in the difference-in-means analysis. The dependent (ordered) variable is farmers' perceptions of soil quality (1=bad, 2=average, 3=good), while factors hypothesized to affect farmers' classification include estimated (AfSIS) soil organic carbon and CEC, maize yield, agricultural inputs, and plot- and household-level characteristics.

In addition, we undertake several descriptive analyses to assess how well the estimated (AfSIS) data capture the soil results found in the measured (soil analysis) data in Kenya. We provide scatter plots to visually explore how the data differ by household. We also report pairwise correlation coefficients and equivalence tests at the village and sample-level to assess whether the AfSIS data can statistically capture the village-level means. Finally we estimate a maize production function with and without soil variables to quantify the "usefulness" of having soil information. We report the results of a Cobb-Douglas[9] production function with three inputs (land, labor, and fertilizer), with and without controls, as well as the predicted yields and marginal physical products of fertilizer for Kenya and Tanzania. Marginal physical product (MPP) measures the additional

[9]For zero values in fertilizer input, we add one to all input levels before taking logs. The point of estimating a production function is to demonstrate whether the coefficients change after including different soil variables. Since returns to inputs may be conditional on soil fertility (Marenya and Barrett 2009b), we also estimate the maize production function with input-soil fertility interaction terms. The coefficients on the interaction terms are not statistically significant.

output that results from the use of one additional unit of input. In our case, MPP of fertilizer measures the additional maize yield in kilograms from using one additional kilogram of fertilizer. By including subjective (farmers' perceptions), measured (soil analysis), and estimated (AfSIS) soil variables in separate specifications, we demonstrate how using different types of available soil information changes estimation and explore whether there is a cost to farmers' or researchers' potential misperceptions.

4 Results and discussion

We present and discuss results for each of our four research questions below.

Question one: farmers' perceptions of soil fertility vs. inputs and yield

Before answering our first research question, we provide three useful descriptive findings that help to shed light on our main results. First, we note the major difference in distribution of plots across farmer-perceived good, average, and bad classifications in western Kenya and Tanzania (Table 1). In Tanzania, only 6 percent of maize plots are classified by their farmers as bad relative to 24 percent in Kenya. In Kenya, over half of all maize plots are regarded as average quality, with a mostly even split of remaining plots between good and bad. In both countries, a majority of farmers classify their soil type as loam (although a higher percentage in Tanzania), with the remaining plots split between clay and sandy.[10]

Second, the correlations between farmer-perceived soil quality and type are also worth mentioning. Farmers distinguish between good and bad soils across all soil types (sandy, loam, and clay) both in Kenya and Tanzania (Table A1 in the Appendix). For example, 15 percent of sandy soils are thought to have good soil quality as opposed to 30 percent of loam soils and 23 percent of clay soils in Kenya. In Tanzania, 43 percent of sandy soils have good soil quality as opposed to 47 percent of loam soils, and 63 percent of clay soils.

Third, it is useful to understand to what extent perceived soil quality measures vary within and across farms; that is, are farmers ranking their fields' fertility relative to others on their own farm? Or do they evaluate their plots relative to the greater context of other farmers in their village or

[10] For the purposes of our work, we drop all plots classified as other from the farmer-perceived soil type analysis. There is only a small percentage of plots in this category in both Kenya and Tanzania.

perhaps further afield? In a decomposition of good, average, and bad perceived soil quality within and between households and villages/enumeration areas (EAs) (Table A2 in the Appendix), we find that variation in plot-level soil quality assessment is largely due to differences between farms within a given village or EA as opposed to within farms.

Table 1 also displays the multiple pairwise comparisons of farmer-reported soil quality and type against agricultural inputs and maize yield levels. We see that the mean values of maize yield are highest on good plots and lowest on bad plots in both Kenya and Tanzania. However, only in Kenya do we find that good plots produce statistically significantly higher yields and only relative to bad plots. Therefore, farmers seem to base their soil quality perceptions on the yield from their maize fields. However, the causal direction of this relationship is not clear from the survey data or our analysis. Loam soils in Kenya have statistically significant higher yield values than do sandy soils.

When looking at inputs used on maize plots, we find that Tanzanian farmers are far more likely to apply some amount of chemical fertilizer on their bad plots than on their good or average ones. This may be an indication that farmers are trying to improve the quality of their bad plots through chemical fertilizer supplements and/or that farmers believe their good or average plots are sufficiently fertile. Average fertilizer application rates (column 4) displayed are conditional on use (column 1). Slightly higher application rates are seen on good and average plots, but the differences are not statistically significant. We find no difference in binary or continuous chemical fertilizer use decisions based on farmer-perceived soil quality in Kenya. We find, however, that loam fields are more likely to receive chemical fertilizer than sandy fields, likely because loamy soils have higher clay and CEC contents and therefore tend to be more responsive to fertilizer use than sandy soils (Lal 2006). We also note that far more farmers in Kenya use chemical fertilizer than do farmers in Tanzania and, therefore, may feel less constrained in their decision to use fertilizer on any of their plots.

With respect to other agricultural inputs, we find that in Kenya good quality plots are more likely to receive herbicide or pesticide than bad plots, but this is not the case in Tanzania. In some parts of Kenya herbicides are often used to prepare land for planting (rather than engaging in more time-consuming human-powered tilling), which could help to explain this finding. Most strikingly, we find that farmers do not vary their organic resources application based on perceptions of soil quality in either Kenya or Tanzania. Only with respect to farmer-reported soil type in Tanzania do we find any statistically significant difference; loam soils are more likely to receive organic resources than clay soils, perhaps because soils high in clay already have relatively high nutrient contents. As organic soil amendements help to rebuild degraded soils, we find it somewhat troubling that farmers do not appear to differentiate their organic resources based on perceived soil quality, especially since most organic resources are generated from on-farm sources (not market-purchased).

Question two: farmers' perceptions vs. objective measures of soil fertility

Table 2 provides results of statistical tests comparing estimated (AfSIS) data to subjective farmers' perceptions of soil quality. Similar to Marenya et al. (2008), we find limited correspondence between farmer-perceived soil quality and estimated (AfSIS) soil data. While plots with good soils have slightly higher organic soil carbon and total nitrogen content and higher pH and CEC than do plots with average or bad soils, the differences between the means are not statistically significant. These soil characteristics, however, vary significantly across the farmer-reported soil types. Soil pH, for example, is lowest (less acidic) on plots with clay soils: 5.72 relative to 5.82 on plots with sandy soils. The pattern is similar for the measurements of soil organic carbon, total nitrogen, pH, and CEC from the soil analysis (AfSIS) data (Table A4 in the Appendix).

We also create an indicator for fertile soils in Kenya based on the three objective measurements and find statistically significant relationships only for soil type. Eighty one percent of plots with fertile soils correspond to plots with farmer-perceived clay soils while only 52 percent of plots with fertile soils correspond to plots with farmer-perceived sandy soils. Soil texture, thus, appears to be the main criteria for soil fertility classification in Kenya.

The picture is somewhat different in Tanzania (Table 2), where we only have objective measurements of soil organic carbon, total nitrogen, pH, and CEC from the AfSIS data. While farmer-perceived soil type remains the main predictor of the differences in objective measurements, plots with better soil quality, as reported by farmers, also have statistically significantly higher carbon content and CEC. Average soil organic carbon content on plots with good soil quality, for example, is 1.67% (w/w) versus 1.57% (w/w) for plots with average soil quality and 1.46% (w/w) for plots with bad soil quality. As the variability between the means is relatively small (as it is in Kenya), the bigger sample perhaps increases statistical significance.

While there are few statistically significant correlations between farmer-perceived and measured or estimated soil fertility indices, especially in Kenya, there are consistent trends in the mean values. With no exceptions, farmer-perceived bad soil is ranked as the lowest measured value across all metrics. Likewise, farmer-perceived good soil is consistently ranked the highest value among all the individual metrics. Similar trends are also found for soil type. Farmer-perceived sandy soils mostly have the lowest measured soil values and the farmer-perceived clay soil mostly have the highest measured values. The notable exceptions include pH and CEC. This is consistent with the known soil chemical characteristics between sandy and clay soils. Although there are few statistically significant differences in measured indicators between farmer-perceived categories, the differences in the means of the measured or estimated metrics are relatively small. This suggests that there are very few differences in empirical soil properties within the study soil (which can also be seen in Figure 2 for Kenya) and, on average, farmer perceptions of soil fertility are sensitive enough to predict subtle changes in soil chemistry.

Moving to multivariate analysis, Tables 3 and 4 show the coefficients after a sequence of the ordered probit estimations. The first column in each table relates farmers' perception to the estimated soil organic carbon and CEC from the AfSIS data; the subsequent columns add covariates by grouping. In Kenya (Table 3), coefficients on soil organic carbon and CEC are not significant, strengthening the results in Table 2. Farmers' perceptions in Kenya do not statistically correspond to the chemical measurements of soil fertility indicators. Coefficients on maize yield, on the other hand, are positive and statistically significant across all the specifications. This indicates that farmers seem to form perceptions of soil fertility based on maize yield (similar result seen in Table 1). Adding other covariates—first inputs then plot and household characteristics—does not change this result. Farmers are also less likely to apply chemical fertilizer on plots perceived to have higher soil quality, and plots with (farmer reported) soil erosion are predictably perceived to have lower soil quality.[11] Farmers' perceptions or misperceptions of soil quality, therefore, do not look clearly targetable based on observed plot or household characteristics (apart from soil erosion). Very similar results are observed in Table 4, the ordered probit estimation for the Tanzanian sample. The main difference is in the sign and significance of the coefficient on soil organic carbon across

[11] Same results hold if we use soil organic carbon and CEC from the measured soil analysis data instead. The results are reported in Table A6 in the Appendix.

all specifications, consistent with the results in Table 2.

Farmers' perceptions of soil quality, therefore, seem to be strongly associated with soil erosion and, as seen above, farmer-observed yields across the two samples. As indicated above, the direction of the soil quality and yield relationship cannot be determined from the data we have available. Farmers may observe yields and conclude that their plots have bad, average, or good soil quality or they may know that their plots have a certain soil quality and then appropriately adjust inputs and other practices to maximize expected yields. This indeterminate causal relationship poses endogeneity concerns for the estimation of agricultural production or profit functions. Therefore, absent more information or an exogenous instrument, one must exercise caution when including farmers' perceptions of soil quality in the estimation of yields.

Question three: high resolution, publicly available soil data vs. researcher-collected plot-level soil data

We find significant differences at the household, village, and sample levels between the two soil data sets for Kenya: the estimated AfSIS soil data and the measured soil analysis data (Table 5).[12] By construction, the AfSIS data show less variation than the soil analysis data, they also suggest different summary statistics than the soil analysis data. While most coefficients are statistically significant ($P<0.05$), they are only high for the two stable indicators of soil fertility (0.68 for soil pH and 0.55 for soil CEC). The two indicators that can vary over time due to both exogenous variables and endogenous management decisions, organic carbon and total nitrogen, have much lower correlation coefficients.

The correlation pattern is also readily observed graphically (Figure 2). The AfSIS data track the soil analysis data, with the soil analysis data showing more variation overall. However, the differences are significant enough to reject most (52 of 64) t-tests of the equivalence of means between the two data sets at the full sample and village levels (Table 6). The notable exceptions are again the more stable soil fertility indicators—soil pH and CEC—where, in each case, 4 of 16 t-tests show that the equivalence of means cannot be rejected. The differences observed across the two samples in terms of average soil organic carbon and total nitrogen content at both the village

[12] We also calculate correlation coefficients between soil quality indicators within the AfSIS data in Tanzania, results for which are found in Table A5 in the Appendix.

and sample level may be partially explained by the differences in sampling periods, as these soil characteristics are subject to change. However, pH and CEC, more stable indicators of soil fertility, are also different for 13 and 12 (out of 15), respectively, villages in the survey and across the full sample for soil pH. The only metric not statistically distinguishable between data sets at the sample level is CEC.

The AfSIS and soil analysis data do, however, exhibit similar patterns when broken down by subjective soil fertility measures (Tables 2 and A4). We find in both the AfSIS and soil analysis data statistically significant differences in soil type (texture) by soil chemistry. However, in moving from the general pattern to the details of the analysis, we again find serious differences between the AfSIS and soil analysis data. In particular, the difference in CEC by soil texture is not observed in the AfSIS data, and the statistically significant discernments of soil texture by soil chemistry differ between the two data sets.

We conclude that these statistically sufficient differences at the household, village, and sample levels justify collection of plot-level soil laboratory analysis data despite the availability of AfSIS data when precise plot-level soil data are important for the analysis at hand (e.g., providing context-specific recommendations to farmers). However, we note that our findings in no way undermine the immense quality and value of the AfSIS data for evaluating landscape-scale soil assessments.

Question four: value of soil information

Finally, and perhaps most important, we consider the value of having specific and accurate soil fertility information. Is soil fertility information necessary to the research that informs policy and programmatic interventions aimed at increasing agricultural productivity and breaking the cycle of rural poverty?

The type of microeconomic analysis that most typically includes soil data is the estimation of production or profit functions. We therefore estimate a series of production functions, starting with specifications that contain no soil information then swapping in the three soil data information types available to us. Tables 7 and 8 show the results of estimating Cobb-Douglas maize production functions for Kenya and Tanzania, respectively. The first two columns of the two tables show estimated coefficients of specifications without and with control variables, in the absence of any soil information. The subsequent columns represent the same basic model but add soil information:

first subjective soil data (farmers' perceptions), then measured data (soil organic carbon and CEC from the soil analysis data in Kenya) and then estimated data (soil organic carbon and CEC from the AfSIS data). We assess the value of the soil information in three ways: (1) changes in the magnitude of coefficients when soil variables are included, (2) changes in predicted maize yield, and (3) changes in predicted marginal physical products (MPP) of fertilizer.

The coefficients on three input variables (land, labor, and fertilizer) are positive, statistically significant, and stable across all specifications. The relative magnitudes and significance levels of the control variables are similarly unchanged. The addition of soil variables does not alter the magnitude of the coefficients on these input variables considerably or increase the models' fit, as represented by the pseudo R-squared values. Predicted maize yields using the estimated coefficients and calculated marginal physical products of fertilizer, conditional on use, are reported in the bottom two rows for all specifications. In neither Kenya nor Tanzania does including soil information of any type change the mean of predicted yields or the magnitude of the standard deviations. A very similar story holds for the estimated average MPP of fertilizer across both samples. Differences in estimated MPP values are better observed in Figure 3, where the distribution within Kenya is explored for each model. The left panel shows the MPPs across the four specifications with control variables at the maize plot level across the full sample, while the right panel zooms in on three villages in the Mid Yala region. For most households, the plot-specific MPP is nearly the same across the models; for some households, measured (soil analysis) soil variables increase the MPP of fertilizer, while estimated (AfSIS) soil variables decrease the value of MPP.

These results perhaps are not surprising. Estimation of the Cobb-Douglas production function offers regression to the mean. While most soil fertility indicators (perceived, measured, and estimated) are positive and statistically significant for both Kenya and Tanzania, they are small in magnitude and correspond to soils with low empirical variation. Therefore, the addition of any soil variables is unlikely to result in vast differences in estimates derived from the underlying models, at least with the methods currently employed and when analyzing in similar contexts (good soils, low empirical variation, and when prediction focuses on sample averages).

The caveats to this discussion are considerable, however. We pursue only one functional form of the production function, do not observe variation across time, and are unable to control for unobserved household or plot-level heterogeneity that could bias our estimates. It is, therefore, conceivable that we could find statistically significant differences in estimated coefficients, yields, or MPPs under somewhat different circumstances. Therefore, when the focus is on specific plots or households, and when heterogeneity across these observations is large, having detailed and accurate soil data still matters. Using data from western Kenya that display a greater degree of soil fertility variation[13] Marenya and Barrett (2009b), for example, find that crop production functions can exhibit von Liebig-type responses. Maize yield response to nitrogen fertilizer in their sample depends on the state of soil fertility, and below some threshold the input applications are not profitable.

While the soil data did not significantly add to the accuracy of estimating maize production functions in our two data sets, it does not suggest that there is not a place for fine-grained soil data in agricultural research. The fine-grained detail provided in plot-level soil analysis data is needed to perform mechanistic and processes-based research at the plot or individual farm scales, whereas high spatial resolution estimated soil data, such as that provided by AfSIS, may be sufficient to meet the needs of those interested in production functions on the country-wide or regional scales. And while farmers' perceptions or misperceptions of soil fertility may not alter the conclusions of a production function analysis, this information can be incredibly informative to extension efforts that seek to identify and correct information gaps.

5 Conclusion

With a renewed appreciation for soil fertility in the international development community, particularly in Sub-Saharan Africa, this paper aimed to take stock of the soil information data currently available to researchers and their analytical limitations. In summary, we find that farmers' perceptions of soil fertility are more correlated with maize yields than with agricultural inputs. We do uncover clear, though not statistically significant, patterns between subjective soil quality assessments and objective scientific measures of soil fertility from laboratory soil analysis and AfSIS data. Although the estimated AfSIS data track the locally collected soil analysis data in Kenya, there are statistically sufficient differences between the two data sets to merit collection of soil samples for the purpose of household and village-level analysis. At the same time, the value of having any

[13] Farms in the data set are sampled based on plot age (time since conversion from forest to agriculture) to capture cultivation time and, therefore, the degree of soil fertility degradation.

soil information in the estimation of simple agricultural production functions appears limited when focusing on average yields and when analyzing in similar context (good soils with low empirical variation).

Overall, we conclude that researchers should not replace objectively observed measures of soil fertility with farmers' perceptions, often collected with household survey data, if one seeks relatively accurate local soil fertility measures and is concerned with the analysis that goes beyond estimation of average yields and returns. But we should not stop asking these questions of farmers either. Many questions remain regarding the role of soil fertility and farmers' response to true or supposed soil fertility. Our analysis has also only considered cross-sectional evidence and described statistical associations. However, given that farmers' perceptions can be learned and that even objective measures vary over time, particularly when and where nutrients are not added back to the soil, a dynamic analysis of any of these correlations could provide even more utility to our disciplines.

As some results from Kenya and Tanzania diverge with respect to farmers' perceptions, the questions remain regarding how farmers make judgments about their soil fertility. We cannot rely on single experiment data or small samples. We need a major research effort to understand how farmers value and use soil information. There is continued need for survey modules that dig deeper into how subjective soil fertility perceptions are formed. Investments should be simultaneously made in (1) understanding the actual learning process farmers use to arrive at their soil fertility distinctions and (2) educating/informing farmers about soil fertility and helping them make their input and other management decisions using this knowledge. As time goes on, we hope to see a better convergence of farmer knowledge with objective soil fertility metrics.

Additionally, questions remain on whether information on soil fertility (such as a soil chemistry information treatment) would alter farmers' behavior in terms of inputs and cropping decisions. In other words, is soil information a limiting constraint to farm management in Sub-Saharan Africa? Investigation into such a question will also enable us to study what farmers do with soil knowledge—does it help improve their farm decisions and, ultimately, yields and welfare measures? Or are farm management decisions informed via some other process? Experimental or quasi-experimental studies could help identify the causal linkages between farmers' perceptions, their management practices, and actual soil fertility to start addressing the soil and human poverty dynamics.

References

Antle, John M., Jetse J. Stoorvogel, and Roberto O. Valdivia. 2006. "Multiple equilibria, soil conservation investments, and the resilience of agricultural systems." *Environment and Development Economics*, 11 477–492.

Barrett, Christopher B., and Leah E. M. Bevis. 2015. "The self-reinforcing feedback between low soil fertility and chronic poverty." *Nature Geoscience*, 8(12): 907–912.

Berazneva, Julia, David R. Lee, Frank Place, and George Jakubson. 2016. "Allocation and valuation of non-marketed crop residues in smallholder agriculture: The case of maize residues in western Kenya." Working Paper, Cornell University, Ithaca, NY.

Bui, Elisabeth. 2010. "Assessment of Digital Soil Mapping products: Independent ground-truthing is essential." In *Proceedings of the 19th World Congress of Soil Science: Soil solutions for a changing world, Brisbane, Australia, 1-6 August 2010. Working Group 1.3 Digital soil assessment.* Institut fur Bodenforschung, Universitat fur Bodenkultur: International Union of Soil Sciences (IUSS), , 9th edition.

Cozzolino, D., and A. Moron. 2003. "The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics." *The Journal of Agricultural Science*, 140(01): 65–71.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review*, 98(2): 482–488.

Guerena, David T., Joseph Kimetu, Susan Riha, Henry Neufeldt, and Johannes Lehmann. 2016. "Maize productivity dynamics in response to mineral nutrient additions and legacy organic soil inputs of contrasting quality." *Field Crops Research*.

Hengl, Tomislav, Gerard B. M. Heuvelink, Bas Kempen, Johan G. B. Leenaars, Markus G. Walsh, Keith D. Shepherd, Andrew Sila, Robert A. MacMillan, Jorge Mendes de Jesus, Lulseged Tamene, and Jrme E. Tondoh. 2015. "Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions." *PLoS ONE*, 10(6): 1–26.

Karltun, E., M. Lemenih, and M. Tolera. 2013. "Comparing farmers' perception of soil fertility change with soil properties and crop performance in Beseku, Ethiopia." *Land Degradation & Development*, 24(3): 228–235.

Lal, Rattan. 2006. "Enhancing crop yields in the developing countries through restoration of the soil organic carbon pool in agricultural lands." *Land Degradation & Development*, 17(2): 197–209.

Lal, Rattan. 2012. "Climate Change and Soil Degradation Mitigation by Sustainable Management of Soils and Other Natural Resources." *Agricultural Research*, 1(3): 199–212.

Lehmann, Johannes, and Markus Kleber. 2015. "The contentious nature of soil organic matter." *Nature*, 528(7580): 60–68.

Marenya, Paswel, Christopher B. Barrett, and Trudi Gulick. 2008. "Farmers Perceptions of Soil Fertility and Fertilizer Yield Response in Kenya." SSRN Scholarly Paper ID 1845546, Social Science Research Network, Rochester, NY.

Marenya, Paswel P., and Christopher B. Barrett. 2009a. "Soil quality and fertilizer use rates among smallholder farmers in western Kenya." *Agricultural Economics*, 40(5): 561–572.

Marenya, Paswel P., and Christopher B. Barrett. 2009b. "State-conditional fertilizer yield response on western Kenyan farms." *American Journal of Agricultural Economics*, 91(4): 991–1006.

Moebius-Clune, B.N., H.M. van Es, O.J. Idowu, R.R. Schindelbeck, J.M. Kimetu, S. Ngoze, J. Lehmann, and J.M. Kinyangi. 2011. "Long-term soil quality degradation along a cultivation chronosequence in western Kenya." *Agriculture, Ecosystems & Environment*, 141(12): 86–99.

Mukhwana, E.J., and J. Odera. 2009. "Best-bet comparison of soil fertility management recommendations in western Kenya: Independent testing by NGOs." 10th KARI Biennial Scientific Conference Volume I, Kenya Agricultural Research Institute (KARI), Nairobi, Kenya.

Niemeijer, David, and Valentina Mazzucato. 2003. "Moving beyond indigenous soil taxonomies: Local theories of soils for sustainable development." *Geoderma*, 111(3): 403–424.

Palacios-Lopez, A., and I.Y. Djima. 2014. "Land Areas in the LSMS-ISA country datasets: An Application of Multiple Imputation (MI) methods." LSMS-ISA Paper, The World Bank (WB), Washington, D.C.

Sanchez, P.A. 2002. "Soil fertility and hunger in Africa." *Science*, 295(5562): 2019–2020.

Sanchez, Pedro A., and M. S. Swaminathan. 2005. "Hunger in Africa: the link between unhealthy people and unhealthy soils." *The Lancet*, 365(9457): 442–4.

Sheahan, Megan, and Christopher B. Barrett. 2014. "Understanding the agricultural input landscape in Sub-Saharan Africa: recent plot, household, and community-level evidence." Policy Research Working Paper WPS7014, The World Bank (WB), Washington, D.C.

Sheahan, Megan, Roy Black, and T.S. Jayne. 2013. "Are Kenyan farmers under-utilizing fertilizer? Implications for input intensification strategies and research." *Food Policy*, 41 39–52.

Shepherd, Keath D., and Meredith J. Soule. 1998. "Soil fertility management in west Kenya: dynamic simulation of productivity, profitability and sustainability at different resource endowment levels." *Agriculture, Ecosystems & Environment*, 71(1-3): 131–145.

Shepherd, Keath D., and Markus G. Walsh. 2002. "Development of reflectance spectral libraries for characterization of soil properties." *Soil Science Society of America Journal*, 66(3): 988–998.

Shepherd, Keith D., and Markus G. Walsh. 2007. "Infrared spectroscopy – enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries." *Journal of Near Infrared Spectroscopy*, 15(1): 1–19.

Sherlund, Shane M., Christopher B. Barrett, and Akinwumi A. Adesina. 2002. "Smallholder technical efficiency controlling for environmental production conditions." *Journal of Development Economics*, 69(1): 85–101.

Sparks, Donald L. ed. 1996. *Methods of soil analysis. Part 3, Chemical methods.* Madison, WI: Soil Science Society of America, American Society of Agronomy.

Stephens, Emma C., Charles F. Nicholson, Douglas R. Brown, David Parsons, Christopher B. Barrett, Johannes Lehmann, David Mbugua, Solomon Ngoze, Alice N. Pell, and Susan J. Riha. 2012. "Modeling the impact of natural resource-based poverty traps on food security in Kenya: The Crops, Livestock and Soils in Smallholder Economic Systems (CLASSES) model." *Food Security*, 4(3): 423–439.

Tittonell, P., B. Vanlauwe, P.A. Leffelaar, E.C. Rowe, and K.E. Giller. 2005a. "Exploring diversity in soil fertility management of smallholder farms in western Kenya: I. Heterogeneity at region and farm scale." *Agriculture, Ecosystems & Environment*, 110(3-4): 149–165.

Tittonell, P., B. Vanlauwe, P.A. Leffelaar, K.D. Shepherd, and K.E. Giller. 2005b. "Exploring diversity in soil fertility management of smallholder farms in western Kenya: II. Within-farm variability in resource allocation, nutrient flows and soil fertility status." *Agriculture, Ecosystems & Environment*, 110(3-4): 166–184.

Tully, Katherine, Clare Sullivan, Ray Weil, and Pedro Sanchez. 2015. "The state of soil degradation in Sub-Saharan Africa: Baselines, trajectories, and solutions." *Sustainability*, 7(6): 6523–6552.

Vagen, Tor-Gunnar, Keath D. Shepherd, Markus G. Walsh, Leigh Winowiecki, Lulseged Tamene Desta, and Jerome E. Tondoh. 2010. *AfSIS Technical Specifications: Soil Health Surveillance.*: Africa Soil Information Service (AfSIS).

Vanlauwe, B., J. Six, N. Sanginga, and A. A. Adesina. 2015. "Soil fertility decline at the base of rural poverty in sub-Saharan Africa." *Nature Plants*, 1(7): , p. 15101.

Vanlauwe, Bernard, Job Kihara, Pauline Chivenge, Pieter Pypers, Ric Coe, and Johan Six. 2011. "Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management." *Plant and Soil*, 339(1-2): 35–50.

Figures and Tables

Figure 1: AfSIS soil pH with the Kenyan soil analysis study households represented by circles. X and Y-axes are latitude and longitude in UTM WGS84.

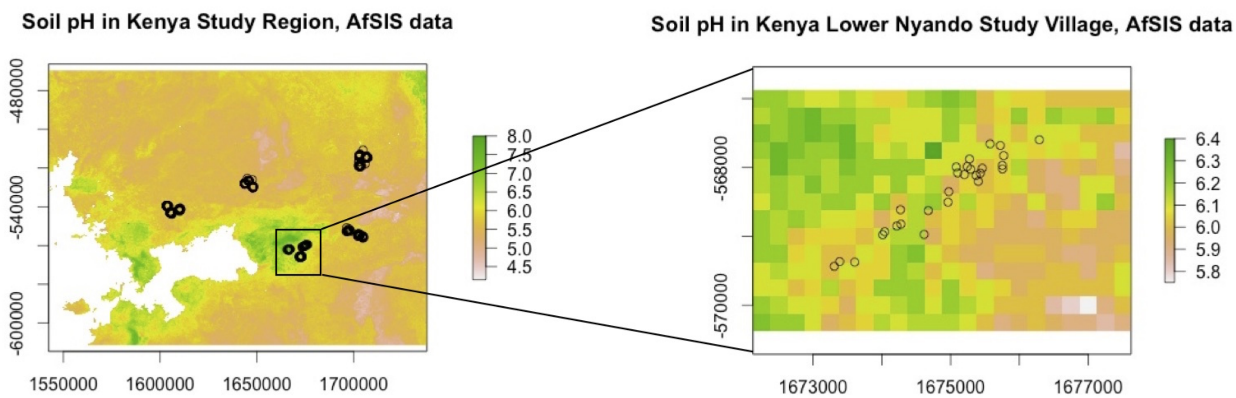

Figure 2: Question 3: Measured (soil analysis) vs. estimated (AfSIS) data by household across the four soil characteristics in Kenya: organic carbon, total nitrogen, pH, and CEC.

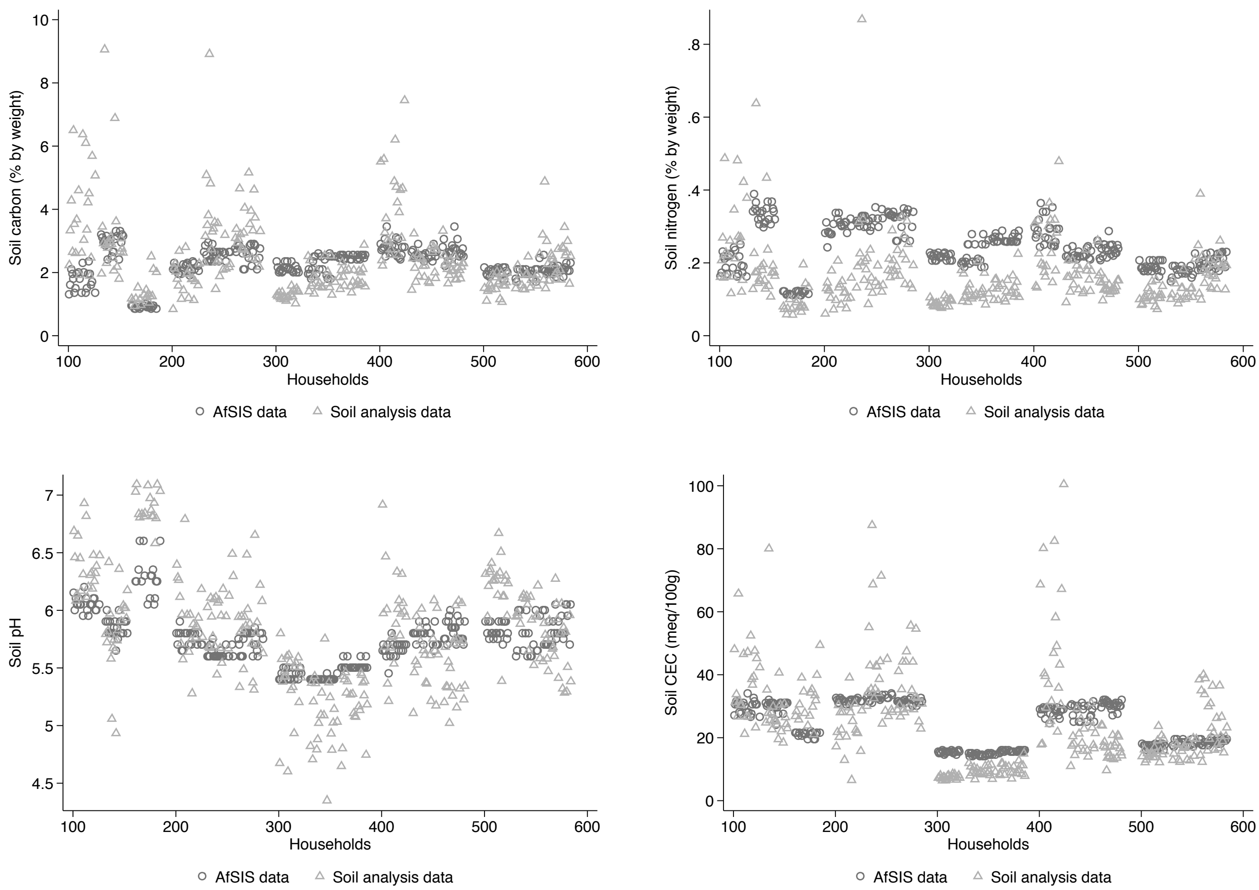

Figure 3: Question 4: Increase in maize yield in response to one additional kilogram of fertilizer (estimated MPP of fertilizer) after the Cobb-Douglas production function with no soil variables, subjective (farmers' perceptions), measured (soil analysis), and estimated (AfSIS) soil variables for all households in the Kenya sample and for households in the three villages in Mid Yala.

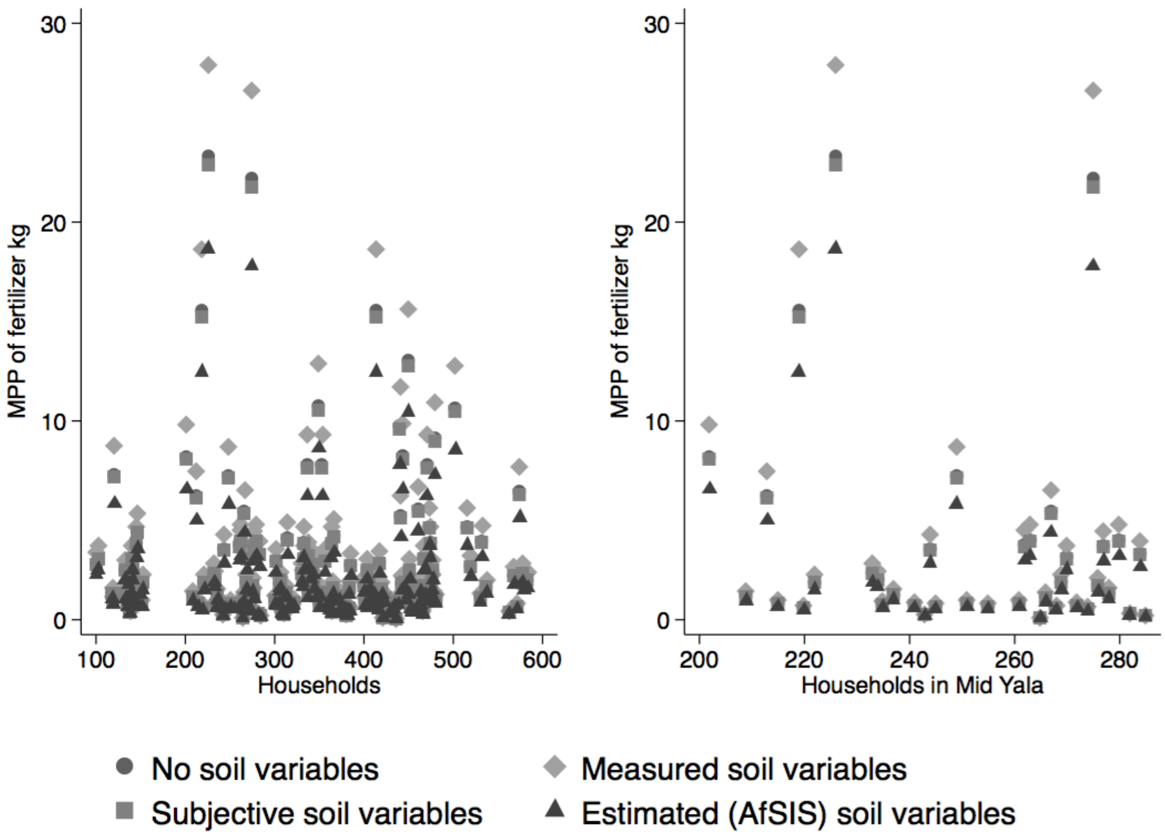

Table 1: Question 1: Subjective soil data (farmers' perceptions) vs. inputs, yield.

Variable	Chemical fertilizer 1=yes	Herbicides, pesticides 1=yes	Organic resources 1=yes	Conditional fertilizer kg/ha	Maize yield t/ha
KENYA					
Soil quality, mean (st.dev.)					
Good (n=124)	0.50 (0.50)a	0.19 (0.40)a	0.64 (0.48)a	137.97 (113.24)a	2.07 (1.70)a
Average (n=262)	0.56 (0.50)a	0.14 (0.34)ab	0.66 (0.48)a	144.08 (136.84)a	1.73 (1.51)ab
Bad (n=123)	0.55 (0.50)a	0.08 (0.27)b	0.67 (0.47)a	120.37 (127.59)a	1.38 (1.30)b
Soil type, mean (st.dev.)					
Clay (n=88)	0.57 (0.50)ab	0.17 (0.38)a	0.65 (0.48)a	149.49 (119.37)a	1.85 (1.41)ab
Loam (n=283)	0.60 (0.49)a	0.14 (0.35)a	0.67 (0.47)a	128.07 (125.15)a	1.83 (1.60)a
Sandy (n=124)	0.42 (0.50)b	0.10 (0.30)a	0.64 (0.48)a	149.45 (154.40)a	1.44 (1.45)b
TANZANIA					
Soil quality, mean (st.dev.)					
Good (n=1152)	0.17 (0.38)b	0.09 (0.29)a	0.15 (0.36)a	146.90 (158.32)a	1.18 (1.35)a
Average (n=1050)	0.18 (0.38)b	0.09 (0.29)a	0.14 (0.35)a	146.29 (143.73)a	1.11 (1.35)a
Bad (n=158)	0.26 (0.44)a	0.10 (0.30)a	0.15 (0.35)a	97.04 (96.78) a	0.94 (1.19)a
Soil type, mean (st.dev.)					
Clay (n=379)	0.21 (0.40)a	0.10 (0.30)a	0.10 (0.31)a	129.90 (112.93)a	1.10 (1.34)a
Loam (n=1536)	0.17 (0.38)a	0.10 (0.29)a	0.15 (0.36)b	147.11 (160.79)a	1.15 (1.33)a
Sandy (n=422)	0.20 (0.40)a	0.07 (0.25)a	0.16 (0.37)ab	133.46 (127.32)a	1.01 (1.34)a

Common letters indicate values are not statistically different at the 95% confidence level using a Tukey-Kramer test.

Table 2: Question 2: Subjective (farmers' perceptions) vs. estimated (AfSIS) soil data.

Variable	Organic carbon (% w/w)	Total nitrogen (% w/w)	pH 1–7	CEC (meq/100g)	Fertile soil**=1
KENYA					
Soil quality, mean (st.dev.)					
Good (n=67)	2.24 (0.52)a	0.25 (0.06)a	5.74 (0.26)a	24.42 (7.23)a	0.75 (0.44)a
Average (n=173)	2.30 (0.52)a	0.24 (0.06)a	5.75 (0.21)a	24.49 (6.79)a	0.73 (0.44)a
Bad (n=68)	2.27 (0.48)a	0.24 (0.06)a	5.78 (0.23)a	23.35 (6.88)a	0.66 (0.48)a
Soil type, mean (st.dev.)					
Clay (n=57)	2.34 (0.51)a	0.25 (0.06)ab	5.72 (0.21)b	25.79 (6.55)a	0.81 (0.40)a
Loam (n=166)	2.33 (0.43)a	0.25 (0.06)b	5.73 (0.23)b	23.63 (7.10)a	0.79 (0.41)a
Sandy (n=75)	2.12 (0.63)b	0.23 (0.07)a	5.82 (0.24)a	24.17 (6.73)a	0.52 (0.50)b
TANZANIA					
Soil quality, mean (st.dev.)					
Good (n=1152)	1.67 (0.92)a	0.12 (0.06)a	6.12 (0.41)a	14.50 (6.68)a	
Average (n=1050)	1.57 (0.90)b	0.12 (0.07)a	6.12 (0.44)a	14.32 (6.72)a	
Bad (n=158)	1.46 (0.75)b	0.12 (0.08)a	6.06 (0.36)a	12.69 (5.63)b	
Soil type, mean (st.dev.)					
Clay (n=379)	1.80 (0.97)a	0.13 (0.07)a	6.08 (0.38)b	14.48 (6.43)a	
Loam (n=1536)	1.62 (0.92)b	0.12 (0.07)a	6.15 (0.45)a	14.74 (6.88)a	
Sandy (n=422)	1.36 (0.71)c	0.11 (0.07)b	6.04 (0.33)b	12.11 (5.35)b	

**Fertile soil in Kenya is defined as soil with C≥2, N≥0.2, and pH≥5.2.
Common letters indicate values are not statistically different at the 95% confidence level using a Tukey-Kramer test.

Table 3: Question 2: Factors affecting farmers' soil fertility perceptions in Kenya.

VARIABLES	(1) Parsimonious	(2) Yield	(3) Inputs	(4) Plot characteristics	(5) Household characteristics
AfSIS soil organic carbon (% w/w)	-0.112	-0.199	0.0202	-0.0798	-0.104
	(0.139)	(0.142)	(0.163)	(0.184)	(0.187)
AfSIS soil CEC (meq/100g)	0.0123	0.00684	0.00601	0.0107	0.0109
	(0.0101)	(0.0103)	(0.0133)	(0.0140)	(0.0145)
Maize grain yield (kg/ha)		0.000179***	0.000206***	0.000180***	0.000179***
		(4.86e-05)	(5.11e-05)	(5.48e-05)	(5.51e-05)
Chemical fertilizer: 1=yes			-0.470***	-0.467***	-0.472***
			(0.167)	(0.176)	(0.177)
Herbicides, pesticides: 1=yes			0.192	0.202	0.212
			(0.203)	(0.219)	(0.221)
Organic resources: 1=yes			-0.0532	-0.119	-0.159
			(0.140)	(0.145)	(0.148)
Improved seeds: 1=yes			-0.0818	-0.106	-0.151
			(0.195)	(0.198)	(0.202)
Maize plot (ha)				-0.0122	-0.0172
				(0.0426)	(0.0445)
Own plot: 1=yes				0.547*	0.584*
				(0.323)	(0.326)
Soil erosion: 1=yes				-0.290**	-0.283*
				(0.144)	(0.145)
Slope: 1=gentle				0.0519	0.0659
				(0.141)	(0.142)
Slope: 1=steep				0.211	0.191
				(0.467)	(0.474)
Distance from home (m)				5.01e-05	4.09e-05
				(0.000169)	(0.000171)
Plot altitude (m)				0.000194	0.000198
				(0.000293)	(0.000301)
Intercropped: 1=yes				0.0147	0.0202
				(0.163)	(0.165)
Household head female: 1=yes					-0.118
					(0.193)
Household head age					-0.00186
					(0.00482)
Household head years of education					0.00566
					(0.0170)
Household size					0.0230
					(0.0296)
Herd size (TLU)					0.00620
					(0.0272)
Observations	307	307	307	307	307
LR χ^2	1.61	15.48	25.60	32.47	34.99
Prob > χ^2	0.447	0.001	0.001	0.006	0.020
Pseudo R^2	0.003	0.026	0.042	0.053	0.057

Ordered probit (coefficients). Dependent variable = Perceived soil quality (1=bad, 2=average, 3=good).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Estimation includes only plots with measured soil data.

Table 4: Question 2: Factors affecting farmers' soil fertility perceptions in Tanzania.

VARIABLES	(1) Parsimonious	(2) Yield	(3) Inputs	(4) Plot characteristics	(5) Household characteristics
AfSIS soil organic carbon (% w/w)	0.0788***	0.0717**	0.0789***	0.101***	0.107***
	(0.0295)	(0.0297)	(0.0300)	(0.0323)	(0.0327)
AfSIS soil CEC (meq/100g)	0.00398	0.00353	0.00125	0.00297	0.00231
	(0.00399)	(0.00399)	(0.00409)	(0.00415)	(0.00421)
Maize grain yield (kg/ha)		0.0000870**	0.000101***	0.000135***	0.000138***
		(0.0000352)	(0.0000360)	(0.0000374)	(0.0000379)
Chemical fertilizer: 1=yes			-0.165**	-0.0883	-0.0901
			(0.0669)	(0.0686)	(0.0696)
Herbicides, pesticides: 1=yes			0.00282	0.0476	0.0418
			(0.0872)	(0.0887)	(0.0894)
Organic resources: 1=yes			0.0110	0.0755	0.0511
			(0.0694)	(0.0709)	(0.0723)
Improved seeds: 1=yes			0.0870	0.0686	0.0924
			(0.0776)	(0.0782)	(0.0800)
Maize plot (ha)				0.0192	0.0193
				(0.0128)	(0.0134)
Own plot: 1=yes				-0.0751	-0.0678
				(0.0764)	(0.0775)
Soil erosion: 1=yes				-0.188***	-0.184***
				(0.0682)	(0.0685)
Slope: 1=gentle				0.0417	0.0497
				(0.0535)	(0.0538)
Slope: 1=steep				-0.108	-0.0976
				(0.128)	(0.129)
Distance from home (m)				0.00000134	0.00000137
				(0.00000114)	(0.00000114)
Plot altitude (m)				-0.000307***	-0.000308***
				(0.0000545)	(0.0000551)
Intercropped: 1=yes				-0.0607	-0.0468
				(0.0503)	(0.0507)
Household head female: 1=yes					0.0916
					(0.0612)
Household head (HH) age					-0.000190
					(0.00183)
HH education: 1=some primary or adult					-0.0268
					(0.0735)
HH education: 1=completed primary					-0.0899
					(0.0681)
HH education: 1=more than primary					0.185
					(0.114)
Household size (adult equivalents)					-0.0145
					(0.0114)
Crop income (USD)					0.0000960
					(0.0000841)
Herd size (TLU)					0.00575
					(0.00438)
Observations	2360	2360	2360	2360	2360
LR χ^2	12.68	18.79	25.69	77.14	91.48
Prob > χ^2	0.002	0.000	0.001	0.000	0.000
Pseudo R^2	0.003	0.005	0.006	0.018	0.022

Ordered probit (coefficients). Dependent variable = Perceived soil quality (1=bad, 2=average, 3=good).
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Question 3: Pairwise correlation coefficients between measured (soil analysis) and estimated (AfSIS) soil data for the four soil characteristics in Kenya: organic carbon (C), total nitrogen (N), pH, and CEC.

		Soil analysis data				AfSIS data			
		C	N	pH	CEC	C	N	pH	CEC
Soil analysis data	C	1.00							
	N	0.96*	1.00						
	pH	0.13	0.07	1.00					
	CEC	0.80*	0.75*	0.43*	1.00				
AfSIS data	C	0.30*	0.23*	-0.48*	0.06	1.00			
	N	0.37*	0.29*	-0.29*	0.25*	0.82*	1.00		
	pH	0.11	0.10	0.68*	0.33*	-0.47*	-0.35*	1.00	
	CEC	0.47*	0.37*	0.26*	0.55*	0.39*	0.58*	0.31*	1.00

*Bonferroni-adjusted significance levels of 0.05 or less. N=307 maize plots.

Table 6: Question 3: Test of equivalence of means between measured (soil analysis) and estimated (AfSIS) soil data in Kenya.

		Organic carbon		Total nitrogen		Soil pH		CEC	
Village	N	t-stat	p-value	t-stat	p-value	t-stat	p-value	t-stat	p-value
Bumira B	21	8.49	0.00	21.77	0.00	6.09	0.00	14.98	0.00
Chamakanga	20	18.68	0.00	45.08	0.00	0.99	0.34	56.53	0.00
Chepkitin B	21	6.41	0.00	13.57	0.00	4.05	0.00	14.61	0.00
Jeveleli	21	3.22	0.00	8.58	0.00	4.74	0.00	18.28	0.00
Kagai	21	-1.82	0.08	2.94	0.01	-5.34	0.00	-2.46	0.02
Kanyibana A	17	-4.38	0.00	3.01	0.01	-10.62	0.00	-6.69	0.00
Kanyilaji B	21	6.35	0.00	14.80	0.00	-7.01	0.00	2.26	0.04
Kasagoma B	21	-3.09	0.01	3.25	0.00	2.36	0.03	-3.52	0.00
Kures	21	-4.36	0.00	8.98	0.00	-1.88	0.08	-1.99	0.06
Lelmolok A	20	3.09	0.01	7.94	0.00	2.35	0.03	6.09	0.00
Nyangera B	21	0.08	0.94	2.67	0.01	-1.96	0.06	0.46	0.65
Ogwedhi B	20	2.72	0.01	18.41	0.00	-2.60	0.02	4.99	0.00
Ratunwet	21	-6.70	0.00	-2.18	0.04	-6.80	0.00	-3.26	0.00
Tabet B	21	-0.70	0.49	5.53	0.00	0.38	0.71	0.36	0.72
Tulwet West	21	-3.78	0.00	1.88	0.07	-3.55	0.00	-3.41	0.00
All villages	308	-2.24	0.03	15.44	0.00	-2.65	0.01	-0.17	0.87

Highlighted values indicate failure to reject statistical difference between soil analysis and AfSIS data.

Table 7: Question 4: Maize Cobb-Douglas production function in Kenya.

VARIABLES	(1) No soil	(2) No soil	(3) Subjective	(4) Subjective	(5) Measured	(6) Measured	(7) Estimated	(8) Estimated
Ln(Land (ha))	0.824***	0.731***	0.810***	0.727***	0.826***	0.743***	0.773***	0.752***
	(0.0792)	(0.0805)	(0.0750)	(0.0782)	(0.0797)	(0.0800)	(0.0806)	(0.0787)
Ln(Labor (days))	0.227**	0.308***	0.231***	0.302***	0.213**	0.289***	0.253***	0.305***
	(0.0889)	(0.0919)	(0.0854)	(0.0895)	(0.0858)	(0.0881)	(0.0894)	(0.0909)
Ln(Fertilizer (kg))	0.199***	0.0861**	0.211***	0.103***	0.183***	0.0845**	0.140***	0.0690*
	(0.0313)	(0.0385)	(0.0297)	(0.0371)	(0.0304)	(0.0380)	(0.0334)	(0.0386)
Plot altitude (m)		0.000654***		0.000601***		0.000519***		0.000516**
		(0.000184)		(0.000179)		(0.000184)		(0.000209)
Herd size (TLU)		0.0335**		0.0302*		0.0319*		0.0376**
		(0.0170)		(0.0176)		(0.0174)		(0.0170)
Intercropped: 1=yes		-0.0328		-0.0174		-0.0402		-0.0686
		(0.120)		(0.117)		(0.119)		(0.123)
Improved seeds: 1=yes		0.389***		0.390***		0.416***		0.428***
		(0.121)		(0.118)		(0.124)		(0.139)
Household head female: 1=yes		0.0458		0.0547		0.0498		0.0734
		(0.126)		(0.125)		(0.124)		(0.129)
Household head age		0.000362		0.000348		0.00181		0.000983
		(0.00321)		(0.00321)		(0.00325)		(0.00331)
Household head years of education		0.0203*		0.0187*		0.0269**		0.0225*
		(0.0112)		(0.0110)		(0.0112)		(0.0120)
Perceived soil quality: 1=average			0.282**	0.157				
			(0.112)	(0.106)				
Perceived soil quality: 1=good			0.584***	0.468***				
			(0.143)	(0.137)				
Measured soil organic carbon (% w/w)					0.263***	0.242***		
					(0.0773)	(0.0741)		
Measured soil CEC (meq/100g)					-0.0170***	-0.0188***		
					(0.00574)	(0.00560)		
AfSIS soil organic carbon (% w/w)							0.334**	0.244
							(0.141)	(0.157)
AfSIS soil CEC (meq/100g)							0.0173**	-0.00937
							(0.00791)	(0.00899)
Constant	4.215***	2.666***	3.896***	2.554***	4.065***	2.690***	3.072***	2.537***
	(0.330)	(0.479)	(0.325)	(0.478)	(0.340)	(0.459)	(0.449)	(0.528)
Observations	307	307	307	307	307	307	307	307
R-squared	0.632	0.680	0.651	0.692	0.651	0.696	0.653	0.684
Maize yield (kg) (in data = 590)	485 (641)	502 (677)	494 (661)	507 (688)	484 (614)	501 (654)	485 (610)	501 (674)
MPP fertilizer (kg) (conditional on use)	5.87 (7.84)	2.54 (3.40)	6.24 (8.32)	3.05 (4.07)	5.40 (7.20)	2.50 (3.33)	4.13 (5.51)	2.04 (2.72)

Dependent variable = Ln(Maize grain (kg)). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The last two rows are means (standard deviations) of predicted maize yields and marginal physical products of fertilizer. Columns 3 and 4 (with and without controls) include subjective soil fertility indicators, columns 5 and 6 include measured soil fertility indicators, and columns 7 and 8 include estimated soil fertility indicators. Estimation includes only plots with measured soil data.

Table 8: Question 4: Maize Cobb-Douglas production function in Tanzania.

VARIABLES	(1) No soil	(2) No soil	(3) Subjective	(4) Subjective	(5) Estimated	(5) Estimated
Ln(Land (ha))	1.458***	1.392***	1.449***	1.384***	1.489***	1.410***
	(0.0467)	(0.0470)	(0.0465)	(0.0468)	(0.0463)	(0.0477)
Ln(Household size (adult equivalents))	0.275***	0.109**	0.279***	0.111**	0.266***	0.100**
	(0.0479)	(0.0487)	(0.0476)	(0.0483)	(0.0460)	(0.0487)
Ln(Fertilizer (kg))	0.142***	0.136***	0.144***	0.136***	0.154***	0.143***
	(0.0124)	(0.0126)	(0.0124)	(0.0126)	(0.0122)	(0.0130)
Plot altitude (m)		0.0000177		0.0000358		-0.00000645
		(0.0000436)		(0.0000440)		(0.0000441)
Herd size (TLU)		0.0234***		0.0229***		0.0235***
		(0.00334)		(0.00336)		(0.00342)
Intercropped: 1=yes		0.103**		0.106***		0.106***
		(0.0410)		(0.0408)		(0.0409)
Improved seeds: 1=yes		0.285***		0.286***		0.254***
		(0.0646)		(0.0642)		(0.0656)
Household head female: 1=yes		-0.127**		-0.128**		-0.130***
		(0.0501)		(0.0500)		(0.0500)
Household head (HH) age		-0.00758***		-0.00751***		-0.00750***
		(0.00146)		(0.00144)		(0.00146)
HH education: 1=some primary or adult		0.0514		0.0543		0.0636
		(0.0570)		(0.0569)		(0.0563)
HH education: 1=completed primary		0.0404		0.0475		0.0506
		(0.0566)		(0.0561)		(0.0561)
HH education: 1=more than primary		-0.107		-0.118		-0.0811
		(0.0922)		(0.0914)		(0.0921)
Perceived soil quality: 1=average			0.128	0.101		
			(0.0844)	(0.0826)		
Perceived soil quality: 1=good			0.258***	0.247***		
			(0.0856)	(0.0841)		
AfSIS soil organic carbon (% w/w)					0.0139	0.0385
					(0.0241)	(0.0261)
AfSIS soil CEC (meq/100g)					0.0140***	0.00985***
					(0.00347)	(0.00374)
Constant	3.728***	4.383***	3.516***	4.162***	3.594***	4.151***
	(0.224)	(0.289)	(0.242)	(0.302)	(0.0904)	(0.269)
Observations	2360	2360	2360	2360	2360	2360
R-squared	0.480	0.513	0.484	0.517	0.444	0.517
Maize yield (kg) (in data = 454)	367 (1,065)	382 (1,199)	368 (1,086)	381 (1,181)	367 (1,149)	385 (1,242)
MPP fertilizer (kg) (conditional on use)	2.02 (4.71)	1.92 (4.49)	2.04 (4.78)	1.93 (4.51)	2.18 (5.09)	2.02 (4.73)

Dependent variable = Ln(Maize grain (kg)). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All regressions include enumeration area level fixed effects to control for unobserved differences in weather and other growing conditions. The last two rows are means (standard deviations) of predicted maize yields and marginal physical products of fertilizer. Columns 3 and 4 (with and without controls) include subjective soil fertility indicators, columns 5 and 6 include estimated soil fertility indicators.

A Appendix: Additional Figures and Tables

Table A1 shows the correlations between farmer-perceived soil quality and soil type. Farmers distinguish between good and bad soils across all soil types both in Kenya and Tanzania. For example, 15 percent of sandy soils are thought to have good soil quality as opposed to 30 percent of loam soils and 23 percent of clay soils in Kenya. In Tanzania, 43 percent of sandy soils have good soil quality as opposed to 47 percent of loam soils and 63 percent of clay soils.

Table A1: Correlations in subjective soil quality and type (farmers' perceptions).

		Soil type			
		Sandy	Loam	Clay	Other
Kenya, 509 plots					
	Good	19	85	20	–
Soil quality	Average	68	136	50	8
	Bad	37	62	18	6
Tanzania, 2,360 plots					
	Good	180	722	239	11
Soil quality	Average	186	732	124	8
	Bad	56	82	16	4

Table A2 shows the variation between good, average, and bad perceived soil quality within and between plots, households, and villages in Kenya and enumeration areas (EAs) in Tanzania.

The first panel of Table A2 indicates the number and percentage of plots that have been designated by their farmers as good, average, or bad in Kenya and Tanzania. In Kenya we see that little over half (51 percent) of the total plots in the sample are perceived as average while there is an even split between good and bad (24 percent each). In Tanzania, nearly half the plots are perceived as good (49 percent) and 44 percent are perceived as average. Only seven percent are perceived as bad. To better understand the source of the variation in perception, the next panels decompose soil quality designation by between and within differences among households and villages/EAs. We observe much greater variation within villages/EAs rather than within households in both Kenya and Tanzania. For example, of the households that report at least one maize plot with good quality in Tanzania, 92 percent of plots within the same household are also deemed to have good soil. On the other hand, of the EAs where someone has declared their soil as good, 57 percent of plots within that same EA have plots with good soil quality. The same applies to the average and bad classifications too.

Table A2: Within vs. between variation in subjective soil quality (farmers' perceptions): Household and village for Kenya and household and enumeration area (EA) for Tanzania.

	Plots		Households			Villages/EAs		
Soil quality	Number	%	Number	% between	% within	Number	% between	% within
Kenya								
312 households, 15 villages								
Good	124	24	98	32	75	15	100	25
Average	262	51	201	64	85	15	100	51
Bad	123	24	91	29	75	15	100	24
Total	509	100	390	125	80	45	400	33
Tanzania								
1,566 households, 292 EAs								
Good	1,152	49	839	54	92	258	88	57
Average	1,050	44	764	49	91	258	88	57
Bad	158	7	126	8	81	93	32	23
Total	2,360	100	1,729	110	91	592	203	49

There are 1.63 maize plots per average household and 33.93 maize plots per average village in Kenya (Berazneva, 2015 data). There are 1.51 maize plots per average household and 8.08 maize plots per average enumeration area (LSMS-ISA data).

Table A3: Number of plots with different farmer-perceived soil quality for 15 villages in Kenya (full sample) and randomly chosen 15 villages in Tanzania.

Kenya	Soil quality			Tanzania	Soil quality		
village	Good	Average	Bad	village	Good	Average	Bad
1	9	6	12	1	4	7	–
2	7	19	8	2	4	3	–
3	8	18	8	3	3	8	–
4	12	12	6	4	1	–	–
5	9	19	5	5	8	–	–
6	8	13	8	6	4	4	–
7	6	17	12	7	6	6	1
8	9	16	4	8	8	3	–
9	10	13	5	9	9	7	1
10	10	17	5	10	4	12	6
11	8	20	5	11	7	6	1
12	7	22	3	12	–	1	1
13	4	28	12	13	–	3	2
14	5	25	12	14	3	7	–
15	12	17	18	15	5	3	–

For Tanzania, 15 villages were chosen using the random number generator at `https://www.random.org/`.

Table A4: Kenya: Subjective (farmers' perceptions) vs. measured (soil analysis) soil data.

Variable	Organic carbon (% w/w)	Total nitrogen (% w/w)	pH 1–7	CEC (meq/100g)	Fertile soil**=1
Soil quality, mean (st.dev.)					
Good (n=67)	2.56 (1.54)a	0.17 (0.12)a	5.85 (0.54)a	25.26 (18.56)a	0.22 (0.42)a
Average (n=173)	2.42 (1.19)a	0.16 (0.08)a	5.81 (0.49)a	24.29 (14.13)a	0.19 (0.39)a
Bad (n=68)	2.32 (0.98)a	0.15 (0.06)a	5.78 (0.54)a	23.59 (14.24)a	0.18 (0.38)a
Soil type, mean (st.dev.)					
Clay (n=57)	2.86 (1.41)b	0.19 (0.09)a	5.90 (0.50)a	30.65 (18.23)b	0.40 (0.49)a
Loam (n=166)	2.34 (1.04)a	0.16 (0.08)ab	5.68 (0.49)b	21.89 (14.28)a	0.16 (0.36)a
Sandy (n=75)	2.27 (1.41)a	0.15 (0.09)b	6.02 (0.50)a	24.23 (12.72)a	0.12 (0.33)b

Common letters indicate values are not statistically different at the 95% confidence level using a Tukey-Kramer test.

Table A5: Pairwise correlation coefficients of estimated (AfSIS) soil data for the four soil characteristics in Tanzania: organic carbon (C), total nitrogen (N), pH, and CEC.

		AfSIS data			
		Carbon	Nitrogen	pH	CEC
AfSIS data	C	1.00			
	N	0.89*	1.00		
	pH	-0.28*	-0.29*	1.00	
	CEC	0.47*	0.49*	0.49*	1.00

*Bonferroni-adjusted significance levels of 0.05 or less.

Table A6 repeats the estimation of the sequence of the ordered probit estimations of Table 6 to explore whether there exists heterogeneity in farmers' perceptions of soil quality in Kenya but using soil organic carbon and CEC from the measured soil analysis data. The dependent variable is farmers' perceptions: 1=bad, 2=average, 3=bad. The first column relates farmers' perception to the estimated soil fertility indicators; the subsequent columns add covariates.

Table A6: Kenya: Factors affecting farmers' soil fertility perceptions (measured soil analysis data).

VARIABLES	(1) Parsimonious	(2) Yield	(3) Inputs	(4) Plot characteristics	(5) Household characteristics
Measured soil organic carbon (% w/w)	0.0878	0.0237	0.0676	0.0500	0.0632
	(0.0863)	(0.0884)	(0.0903)	(0.0925)	(0.0949)
Measured soil CEC (meq/100g)	-0.00295	0.000379	-0.00365	0.000203	-0.000796
	(0.00698)	(0.00706)	(0.00737)	(0.00766)	(0.00771)
Maize grain yield (kg/ha)		0.000166***	0.000201***	0.000176***	0.000174***
		(4.78e-05)	(5.16e-05)	(5.53e-05)	(5.56e-05)
Chemical fertilizer: 1=yes			-0.483***	-0.504***	-0.512***
			(0.152)	(0.166)	(0.169)
Herbicides, pesticides: 1=yes			0.207	0.209	0.224
			(0.202)	(0.219)	(0.222)
Organic resources: 1=yes			-0.0621	-0.116	-0.157
			(0.139)	(0.143)	(0.146)
Improved seeds: 1=yes			-0.0291	-0.0623	-0.106
			(0.161)	(0.172)	(0.178)
Maize plot (ha)				-0.00835	-0.0150
				(0.0416)	(0.0439)
Own plot: 1=yes				0.508	0.530
				(0.321)	(0.323)
Soil erosion: 1=yes				-0.299**	-0.287*
				(0.146)	(0.147)
Slope: 1=gentle				0.0488	0.0613
				(0.140)	(0.141)
Slope: 1=steep				0.268	0.229
				(0.466)	(0.473)
Distance from home (m)				4.96e-05	4.48e-05
				(0.000169)	(0.000170)
Plot altitude (m)				0.000169	0.000151
				(0.000267)	(0.000272)
Intercropped: 1=yes				0.0123	0.0102
				(0.162)	(0.164)
Household head female: 1=yes					-0.101
					(0.192)
Household head age					-0.000878
					(0.00487)
Household head years of education					0.00872
					(0.0173)
Household size					0.0235
					(0.0296)
Herd size (TLU)					0.00829
					(0.0266)
Observations	307	307	307	307	307
LR χ^2	1.45	13.76	25.91	32.62	35.14
Prob > χ^2	0.485	0.003	0.001	0.005	0.019
Pseudo R^2	0.002	0.023	0.043	0.054	0.058

Ordered probit (coefficients). Dependent variable = Perceived soil quality (1=bad, 2=average, 3=good). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.