A Prequel to the Space to Place Impact Evaluation - Assessing the Relative Importance of Space and Place Components

Thomas Asefa,[*] Joachim De Weerdt,[†]
David Spielman,[‡] Bjorn Van Campenhout[§]

July 17, 2024

A pre-analysis plan

1 Introduction

Raising agricultural productivity among smallholder farmers in Sub-Saharan Africa is widely recognized as an important component of inclusive wealth creation and structural transformation. Central to this endeavor will be the adoption of sustainable soil and land management to improve the sustainability, resilience and productivity of agriculture. The USAID Bureau for Resilience and Food Security (RFS) has led a partnership with the International Fertilizer Development Center (IFDC) and other partners to build a low-cost "Space to Place" (S2P) approach for developing and disseminating localized fertilizer guidance. The S2P approach revolves around the delivery of spatially appropriate soil fertility management recommendations, guided by digitized soil maps (Space) combined with farm(er)-level characteristics (Place), for effective agronomic and fertilizer recommendations that increase fertilizer use efficiency and maintain or surpass current productivity levels, and reduce fertilizer wastage.

IFPRI was tasked in 2023 to do a rigorous impact evaluation of the S2P approach. It is now becoming apparent that the S2P prototype will not be ready to be evaluated during the 2024/25 growing season. A full and rigorous evaluation of the S2P approach should be possible for the 2025/26 growing seasons and beyond.

We are proposing to use the 2024/25 growing season to generate learning to inform S2P's theory of change. A unique opportunity presents itself through

[*]Innovation Policy and Scaling Unit, IFPRI, Kenya
[†]Development Strategy and Governance Division, IFPRI, Malawi
[‡]Innovation Policy and Scaling Unit, IFPRI, United States
[§]Innovation Policy and Scaling Unit, IFPRI and LICOS Centre for Institutions and Economic Performance KU Leuven, Belgium - corresponding author: b.vancampenhout@cgiar.org

S2P's partnership with FSU Meridian. This private-sector player offers farmers spectral soil tests and site- and crop-specific recommendations. The model underlying the test was calibrated on wet-chemistry analyses conducted on 22,000 samples from across Malawi and has been rolled out across X years, now reaching Y farmers. The tool is part of a private-sector business model in which FSU Meridian forms a recommendation for a package of nutrients, which is then supplied through their network of 400 agrodealers.

2 Motivation

Decision support tools for tailored nutrient recommendations have generally fallen short of expectations (Sida et al., 2023). This could be due to the quality of the information they provide, and it has indeed been argued that, often, "...the numerous laboratory services and digital applications providing field-specific recommendations appear to promise more accuracy than soil analysis can realistically deliver" (Schut and Giller, 2020). Another reason why such tools have not scaled may be due to the fact that they "address in-field heterogeneity, but not socioeconomic heterogeneity of users" (Sida et al., 2023). Indeed, one of the only rigorous impact evaluations that finds a positive effect of a decision support tool also has an important agronomic information component (Ayalew, Chamberlin, and Newman, 2022; Beg, Islam, and Rahman, 2024; Gars, Kishore, and Ward, 2022). In light of this, we feel it is important to study the relative importance of the Space and Place components in a clean field experiment.

3 Research Questions

1. Does the provision of site-specific soil nutrient information and the recommendation of a specific fertilizer blend based on this information, change outcomes of interest. The motivation for this research question can be found in the "Space" component of S2P.

2. Does the provision farm-specific and farmer-tailored agronomic information with respect to soil nutrient management--and particularly recommendations on how to improve on current practice--change outcomes of interest. This research question speaks to the "Place" component of S2P.

3. Does the provision of site-specific nutrient information and farm-specific and farmer-tailored agronomic information with respect to soil nutrient management change outcomes of interest. This research question focuses on potential interactions between the Space and Place components as implied in the S2P approach.

4. Does the provision of site-specific nutrient recommendations and agronomic information with respect to soil nutrient management, combined with a subsidy to buy these nutrients at a local agro-dealer change the

outcomes of interest. This research question looks at the importance of cash/credit constraints in the effectiveness of S2P.

4 Sample and external validity

The study area requires on the one hand the presence of Meridian's agro-dealers (to ensure availability of the recommended inputs and to run the voucher scheme through), but on the other hand a population that has not yet been exposed to FSU's soil testing and nutrient recommendation system. We intend to work with FSU to identify greenfield locations, where FSU is currently not operational, but where it is considering expansion by setting up new agro-dealerships. We would then work with FSU to adapt their approach in these areas for 2024/25 season only. Within the market shed of the new shop, we will select X villages and compile a list all farmers with at least 0.4 ha of land. From this list we will randomly assign farmers into the 5 different treatment groups.

5 AIP recipients

We are expecting about one out of every four farmers in our sample to be receiving an AIP subsidy for the purchase of 50kg of the standard NPK blend and 50kg of Urea. The prevalence of AIP recipients will be balanced across treatment and control (that is, we will block on being an AIP recipient in some way). This will also allow us to look at treatment heterogeneity of the intervention on this attribute (impact of interventions among AIP recipient versus impact of interventions among non-AIP recipient) - see next section.

6 Impact heterogeneity

We will study heterogeneity across gender, land size, wealth, remoteness, availability of labor in the household, receipt of an AIP subsidy, and use of improved agricultural technology (irrigation, improved seeds, soil health management,. . .).

7 Treatment Protocols

- Treatment (T1): a farmer is offered a free soil test on their main plot. If they accept, a few days later the farmer receives a soil report (see example in Appendix 1), with some key characteristics of the soil. The report indicates the soil's texture, pH level, cation exchange capacity, percentage organic matter, total nitrogen, carbon to nitrogen ratio, as well as the parts per million of available P, exchangeable K, calcium, magnesium, iron, manganese, boron, copper and zinc. Beyond the exact measurements each soil property is given a score on a 4-point Likert scale (very low,

low, high, very high), with emojis, color coding and bar charts indicating where the soil is deficient and where the levels are sound. Using a script, an enumerator then discusses the soil report with the farmer to derive a recommended nutrient package specific to the crop the farmer wants to grow. The farmer is then informed where and how to obtain these nutrients. The philosophy behind this treatment is really to provide objective information about the soil on the plot that was chosen. The enumerator is explicitly told to not comment on or make recommendations with respect to farmer management practices.

- Treatment (T2): using a simple application, data is collected on a set of key soil nutrient management practices that the farmer used in the past, that the farmer generally uses, and/or that the farmer has the intention to use in the next season (fertilizer application rates, timing of application, method of application, visual cues that may suggest issues with soil health,...). This information is then used to come up with tailor-made suggestions on how the farmer can improve. The philosophy behind this treatment is to focus only on practices and readily observable features of the crop and plot. The application will be developed on the basis of a workshop with experts. During this workshop, problem tree analysis will be used to identify current practice, and potential improvements for each problem will be selected. Experts will then rank different solutions based 1) on how common/known each solution is and 2) expected effect size on key outcomes. Based on these insights, decision flowcharts will be developed and implemented in ODK.

- Treatment (T3): a voucher to buy the recommended nutrients at a local agro-dealer. The voucher will be of a value equivalent to the government's fertilizer subsidy.

- Comparison (C): A pure control with no soil test, no recommendations on optimal soil fertility management, and no subsidy.

8 Experimental Setup

Farmers will be randomized at the individual level following a factorial design for T1 and T2. An extra treatment cell will be added to the crossed treatment cell to measure the incremental effect of the subsidy (T3). The layout of the experiment is illustrated in Figure 1. To test the effectiveness of T1, we can exploit the factorial design and pool T1 and T1+T2 treatment cells and compare this to the pooled treatment cells C and T2. Similarly, to estimate the effect of T1, we can pool T2 and T1+T2 and compare this to the pooled C and T1 cells.[1] To test the (additional) effect of T3, we simply compare outcomes of treatment cell

[1] It is well known that the increase in power due to treatment cell pooling can lead to both bias and false positives if there are significant interaction effects. Section 9 explains strategies to deal with this issue.

C	T1	
T2	T1+T2	
		T1+T2+T3

Figure 1: Experimental layout

T3 to outcomes for experimental subjects in treatment cells T1+T2. Another interesting comparison we will consider is the difference between C and T1+T2, which can be thought of as the impact of the S2P intervention as a package.

9 Empirical Strategy

Due to the random assignment of participants to treatment and control groups, comparing outcome variable averages of treated and control participants provides unbiased estimates of the average treatment effects. We compare averages using OLS regression. The exact specification will depend on the hypothesis we want to test.

For simple comparison, such as comparing C to T1+T2, or comparing T1+T2 to T3, we use the following specification:

$$Y_i = \alpha + \beta_T T_j + \varepsilon_i \tag{1}$$

where Y_i is the outcome of interest (agronomic practices, profits, return on investment, etc—see Section 11 below) of farmer i. T_i is a treatment indicator that takes the value of one if the farmer was allocated to the treatment of interest (T1+T2 to estimate the effect of S2P as a package and T3 to estimate the impact of adding a voucher to the package) and zero if it is in the control group of that treatment (C to estimate the effect of S2P as a package and T1+T2 to estimate the impact of adding a voucher to the package)). The parameter estimate of β_T is the parameter of interest.

For the regressions that estimate the effect of the components of S2P separately, we exploit the factorial design to boost power by pooling across treatment cells (See Section 8). However, we do not want to rule out the possibility of an interaction effect between the two components, and as we will see in Section 10, powering our design to detect a reasonable interaction effect proves prohibitively expensive. Therefore, we estimate the following equation:

$$Y_i = \alpha + \beta_T T_j + \beta_O \left(O_j - \bar{O}\right) + \beta_I T_j \left(O_j - \bar{O}\right) + \varepsilon_{ij} \tag{2}$$

where the orthogonal treatment (O_j) enters the equation demeaned and fully interacted with the main treatment (Lin, 2013; Muralidharan, Romero, and

WÃŒthrich, 2023). This approach, that borrows from the literature on covariate adjustments in experiments, allows one to recover the parameter estimate of the fully interacted model without losing the power advantage of factorial designs.[2]

For outcome families with more than one outcome (knowledge, adoption, input use, and effort), we compute outcome indices, which is a common way to account for multiple hypothesis testing. To do so, we follow Anderson (2008), where each index is computed as a weighted mean of the standardized values of the outcome variables. The weights are derived from the (inverse) covariance matrix, such that less weight is given to outcomes that are highly correlated with each other. For these indices, signs of outcomes were switched where necessary so that the positive direction always indicates a "better" outcome.

10 Statistical Power and Sample Size Determination

We will run a simulation analysis to determine sample size for the design presented in section using the specifications in. Simulation, where you basically tell your computer to run an experiment thousands of times and simply count how frequently our experiment comes up significant under a set of assumptions, is a far more flexible, and far more intuitive way to think about power analysis.[3]

We will use yield (kg/ha) as the main outcome in the power analysis. We assume a normal distribution with mean 2,600 kg/ha with a standard deviation of 2,000.

A key parameter in statistical power analysis is the choice of the minimal detectable effect (MDE) size. This can be done by looking at previous studies that use similar treatments. One such treatment is the one used in the study in Ethiopia reported in Ayalew, Chamberlin, and Newman (2022), which consists of providing information on the recommended amount and blend of fertilizer to use on a particular maize plot, the optimal timing of the application of that fertilizer, and the expected yield outcome for that recommendation. They report a 13 percent increase in yield (437 on baseline yield of 3250 kg/ha) and a very similar increase in farmer profit. Another interesting study is Mesfin et al. (2021) who test different approaches in field trials for barley production in Alaje, northern Ethiopia. They have a design with 4 treatment arms. The first is based on the Quantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model with yields up to 4747. The second is a blended fertilizer application (which has an average yield of 4045); the third treatment arm implements average farmers' fertilizer application practices (where yield was 3,637) and there also a control group (without any fertilizer, obtaining 2727 kg/ha). Form this, we estimate a

[2]See also this blog post: https://blogs.worldbank.org/en/impactevaluations/what-should-you-do-experiments-factorial-designs

[3]See also this methods guide: https://egap.org/resource/10-things-to-know-about-statistical-power/

treatment effect that ranges from 11 percent (if all farmers are using following farmer practices) up to 50 percent (if farmers were not using fertilizer before).

We will use a 15 percent increase in yield from T1 and a 10 percent increase in yield from T2. We further model a 7.5 percent interaction effect. Finally, when comparing groups T1*T2 with T3, we expect 12.5 percent increase.

For the outcome variable, which we consider to be yield, we assume a normal distribution with mean 2,600 kg/ha with a standard deviation of 2,000.

We run two experiments. We first power the factorial design, with subsamples of the same size in each of the four treatment cells. We then look at power of four comparisons. First, we compare T1 to the control using Equation 2. Similarly, we compare T1 to the control using the same equation. As these pool two treatment cells, power is likely to be high at a given sample size. We also look at the interaction effect of a fully saturated model:

$$Y_i = \alpha + \beta_{T1} T1_j + \beta_{T2} T2_j + \beta_{T12} T1_j . T2_j + \varepsilon_i \tag{3}$$

As this compares treatment effects in a single cell over and above the main treatment effects, power is likely to be low. We also directly compare the control to the T1+T2 treatment. As expect a large effect size here, the smaller samples used due to the fact that we only compare two treatment cell should be less of an issue. Finally, we also compare the two treatment cells of T1+T2 to T3, which also has a reasonable minimal detectable effect size. Other parameters for the simulation include a 5 percent significance level, and 1000 replications.

Resulting power curves are plotted in Figure 2. The orange curve, representing the comparison of the T1+T2 package to the pure control, has highest power. With a total sample size of just over 300 (meaning less than 100 farmers in each of the four groups), we would get 80 percent power for that test. The red curve compares T2 to the pure control. Here we get 80 percent power with a total sample of about 900 farmers or 225 in each treatment group. The black line is for the T1-C comparison, which has higher power (due to the larger assumed MDE). Finally, the blue line depicts power for the interaction effect. Power for the interaction effects remains minimal over the entire range of sample size we consider reasonable.

Panel (b) of Figure 2 indicates that we would need about 500 to 600 farmers for the comparison between T1+T2 and T3, or 250 to 300 in each group.

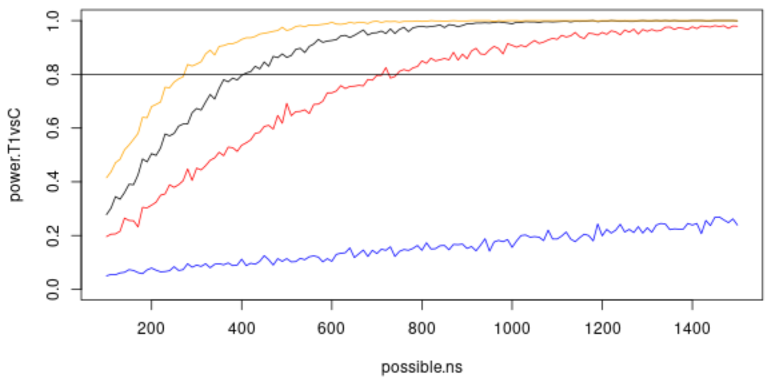
Given the above, we settle for a total sample of 1,250, with 5 groups of 250 farmers.

11 Outcomes of interest

Our outcomes of interest are agricultural practices (with a particular focus on the use of recommended inputs and practices), yields, return on investment in fertilizer, profitability of the targeted crops and the return on investment for the subsidy.

- number of crops grown by the farmers

(a)



(b)

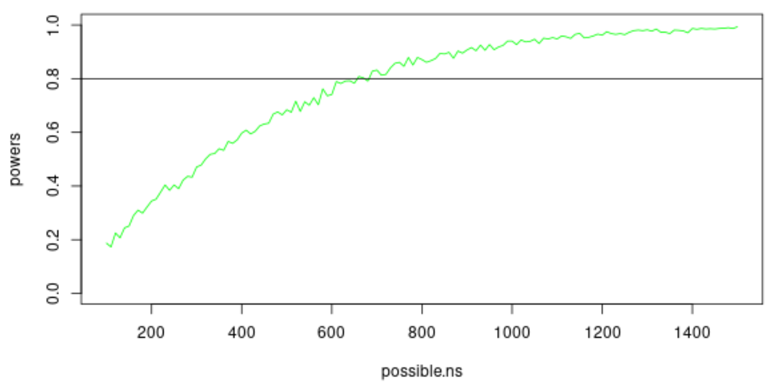


Figure 2: Power curves

- use of inorganic fertilizer on any field
- use of inorganic fertilizer on the main plot

We will also test for balance on ten pre-specified variables. While we will not do a dedicated baseline, we will collect data on some background characteristics at the time of the treatment administration. We will use the following five variables that are unlikely to be affected by the treatments: farmer's age (in years), sex of farmer, household size, land area for crop production (acres), and AIP recepient in previous year. We also include five variables from among the outcomes of interest to test balance at baseline, in particular: number of crops grown by the farmer, use of seed of an improved variety, yield in previous season on the main plot.

12 Timeline

Week of july 15: workshop in Malawi to develop treatments September 2024: soil tests October 2024: treatment administration (+ baseline data collection) April 2025: endline data collection

================================

13 Problems and consideration

A first question is related to the agronomic or the effect size of a fairly light touch intervention. Throughout the scoping, we found it hard to get researchers to quote effects of provide us with yield response curves. The problem is confounded by the fact that yield effects are likely secondary to cost benefit considerations when farmers decide to use a particular blend of fertilizer.

The numerous laboratory services and digital applications providing field-specific recommendations appear to promise more accuracy than soil analysis can realistically deliver (Schut and Giller, 2020).

The site specificness of many of the applications also needs to be taken with a grain of salt. In most cases, the input for the decision support systems consists of things like farmer agro-management practices and soil characteristics as evaluated by the farmers using rudimentary characteristics (clay, sandy, silt, or loam soil type). The models behind the decision support systems are often also not very specific, focusing on broad agro-ecological zones or administrative boundaries like districts. However, there will be plenty of farmers that are on the periphery for which alternative recommendations will be more effective. This is a serious concern, as Beg, Islam, and Rahman (2024) find little correlation between recommendations based on union-councils and actual nutrient content on the plots within these union-councils. More could be done by exploiting the geo-location capabilities of digital decision support tools.

Another problem is that some applications focus on maximizing yields and to not consider costs. However, optimal fertilizer application is likely to change with both input and output prices. Again, the extent to which this can be integrated into an application can vary. The CABI app for instance, does consider the cost of inputs when optimizing fertilizer allocation but farmers have to input the prices of the fertilizers. A better way would be to link the application to prices that get updated in real-time.

Site specific recommendations are not very effective is farmers can not access the blends or components being recommended. This was a problem that was repeatedly mentioned when we discussed site specific recommendations during scoping. Here, in theory, the use of ICT may provide an excellent opportunity to link advise to stocks in real time. For instance, when recommendations are provided, the application may point farmers to the nearest agroworld shop that has the particular blend in stock. Alternatively, the decision support tool may resort to a second best blend if the first choice can not be obtained within a reasonable distance.

Somewhat related, the total cost of the blend should also be considered. It was mentioned that farmers know the optimal blend, but that they will go for the cheaper option. This

Indivisibility, rule-of-thumb etc. Farmers just use what they get, regardless of how much land they have and what soil quality is. Or farmers do not pay attention to quantities. Do not say use 20kg of dap and 10 kg of urea, but use twice as much dap as urea. Again, the idea of some kind of market place for fertilizer would help...

14 Design

In the first year, we will work with a paper based version of the decision support. We will have 4 different treatment arms in the form of a factorial design. In a first treatment cell, we will have site specific recommendations. In a second treatment, we will provide extension on fertilizer application using a principled approach. In a third treatment cell, we combine extension on fertilizer application with extension. A final treatment cell functions as a pure control.

In the second year, the decision support tool will be piloted, but this time with an alternative

15 Related literature

Recently, various studies have been done on site specific fertilizer recommendations.

Ayalew, Chamberlin, and Newman (2022) evaluate the impact of targeted site-specific fertilizer blend recommendations to Ethiopian smallholder farmers on fertilizer usage, farm productivity, profits from maize production, and household welfare using a two-level cluster randomized control trial. The problem with this study is that it not only provides site specific recommendations, but also information on timing on application and on expected yields. However, in the control group, no intervention takes place. A better control group would have received the same intervention with the only difference that the regional recommendations. We feel that the treatment effect is contaminated by an effect emanating from making fertilizer more salient.

Beg, Islam, and Rahman (2024) do not find convincing effects from fertilizer quantity recommendations in Bangladesh. They compare two types of variety-specific fertilizer recommendations – government provided, community-level recommendations and plot-specific recommendations based on individual soil tests. One potential reason why they find little effect (they only find that farmers stop using a particular fertilizer type because farmers overuse according to both recommendation types) is that the interventions are too complex. In a different experiment, Islam and Beg (2021) provide farmers with a simple tool (a leaf color chart) and basic, rule-of-thumb instructions to guide the timing and quantity of urea (nitrogen) application. This intervention with substantially lower mental and attention demands on farmers does have a significant effect on the quantity of urea (nitrogen-fertilizer) applied. Taken together, this led them to conclude that advice should be bundled with training and extension services to ensure farmers benefit as intended.

Berazneva et al. (2023)

Beg and Islam (2020)

Gars, Kishore, and Ward (2022) do not find compelling evidence that the information treatment led farmers to substantially change their overall fertilizer application nor their willingness to pay for recommended micronutrients, though there is some evidence that farmers may have altered the timing of their fertilizer application in such a way that improves fertilizer use efficiency.

Additional treatments

Credit and/or insuracence. Note that Ayalew, Chamberlin, and Newman (2022) provide free insurance as a complementary treatment and find no effect of this.

16 Outcomes

- a dummy for fertilizer use/dummies for use by fertilizer type
- total level of inorganic fertilizer used in kg/ha. Note that this effect is likely to be positive, but there may be some heterogeneity as farmers that use too much fertilizer may actually reduce fertilizer use. Ayalew, Chamberlin, and Newman (2022) therefore also estimate quantile regressions and find a negative effect in higher quantiles.
- Application error or the gap between the amount of fertilizer used and the recommended amounts. This is a key outcome in Ayalew, Chamberlin, and Newman (2022), who define this as follows: "For the treatment households, the fertilizer use gap is calculated by using the absolute difference between the actual macronutrient in kg/ha that farmers applied for maize production and the site-specific recommended values, whereas

for the control house holds we used the absolute deviation of the actual macronutrients used by farmers and the regional recommendation values.". This could also be modeled as a quadratic loss function. In Beg, Islam, and Rahman (2024), the application error is constructed for farmers who apply a non-zero amount of the respective fertilizer and is the absolute difference between endline usage and the recommended quantity for each fertilizer (and it is not clear what recommended quantities they use). They standardize the application error with respect to the control group mean and standard deviation.

- productivity: measured as production of maize per acre.
- revenue: yield*price
- farm expenditure (inputs + labour)
- profits
- welfare

References

Anderson, M. L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Ayalew, H., J. Chamberlin, and C. Newman. 2022. "Site-specific agronomic information and technology adoption: A field experiment from Ethiopia." *Journal of Development Economics* 156: 102788.

Beg, S. and M. Islam. 2020. "Improving Fertilizer Management can lead to Medium Term Increase in Complimentary Inputs." Working Paper.

Beg, S., M. Islam, and K. W. Rahman. 2024. "Information and behavior: Evidence from fertilizer quantity recommendations in Bangladesh." *Journal of Development Economics* 166: 103195.

Berazneva, J., A. Maertens, W. Mhango, and H. Michelson. 2023. "Paying for agricultural information in Malawi: The role of soil heterogeneity." *Journal of Development Economics* 165: 103144.

Gars, J., A. Kishore, and P. Ward. 2022. "Confidence and information usage: Evidence from soil testing in India." *Unpublished Working Paper* .

Islam, M. and S. Beg. 2021. "Rule-of-Thumb Instructions to Improve Fertilizer Management: Experimental Evidence from Bangladesh." *Economic Development and Cultural Change* 70 (1): 237–281.

Lin, W. 2013. “Agnostic notes on regression adjustments to experimental data: Reexamining Freedmanâs critique.” *The Annals of Applied Statistics* 7 (1): 295–318.

Mesfin, S., M. Haile, G. Gebresamuel, A. Zenebe, and A. Gebre. 2021. “Establishment and validation of site specific fertilizer recommendation for increased barley (Hordeum spp.) yield, northern Ethiopia.” *Heliyon* 7 (8).

Muralidharan, K., M. Romero, and K. WÃŒthrich. 2023. *Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments*. Tech. rep.

Schut, A. G. and K. E. Giller. 2020. “Soil-based, field-specific fertilizer recommendations are a pipe-dream.” *Geoderma* 380: 114680.

Sida, T. S., S. Gameda, J. Chamberlin, J. A. Andersson, M. Getnet, L. Woltering, and P. Craufurd. 2023. “Failure to scale in digital agronomy: An analysis of site-specific nutrient management decision-support tools in developing countries.” *Computers and Electronics in Agriculture* 212: 108060.