Using Satellites and Phones to Evaluate and Promote Agricultural Technology Adoption: Evidence from Smallholder Farms in India *

Shawn Cole
Harvard Business School

Grady Killeen
University of California, Berkeley

Tomoko Harigaya
Precision Development

Aparna Krishna
Indian Institute of Technology (Indian School of Mines) Dhanbad

July 2023

Abstract

This paper evaluates a low-cost, customized soil nutrient management advisory service in India. As a methodological contribution, we examines whether and in which settings satellite measurements may be effective at estimating both agricultural yields and treatment effects. The intervention improves self-reported fertilizer management practices, though not enough to measurably affect yields. Satellite measurements calibrated using OLS produce more precise point estimates than farmer-reported data, suggesting power gains. However, linear models, common in the literature, likely produce biased estimates. We propose an alternative procedure, using two-stage least squares. In settings without attrition, this approach obtains lower statistical power than self-reported yields; in settings with differential attrition, it may substantially increase power. We include a "cookbook" and code that should allow other researchers to use remote sensing for yield estimation and program evaluation.

*We gratefully acknowledge financial support from ATAI and HBS' Division of Research and Faculty Development. We thank Claudia Carbajal Morelos and Arnesh Chowdhury for superb research management and support, Jaagruti Didwania, Azfar Karim, and Prathyush Parasuraman for excellent research assistance, Tarun Pokiya for his agricultural expertise and suggestions, and the Precision Development (PxD) India team for implementation support. Disclosure: Shawn Cole is a board member of Precision Development (PxD), an NGO delivering mobile phone-based extension to small-holder farmers. Cole does not receive any financial compensation from the organization. Harigaya is, and Killeen and Krishna were, employed by PxD. Cole: scole@hbs.edu; Harigaya: tharigaya@precisiondev.org; Killeen: gkilleen@berkeley.edu, Krishna: apakrishna@gmail.com

1 Introduction

Information technology has transformed farming in the developed world. Precision farming - farm management informed by increasingly data-driven, dynamic, and accurate instructions - has been credited with raising profitability by as much as 3% (Schimmelpfennig, 2016). The dramatic reduction in costs of collecting and disseminating information in developing countries offers the tantalizing prospect that such approaches may improve productivity among small-holder farmers as well. This paper evaluates how such an approach might be adapted in a low-income setting, testing whether mobile phones can be used to deliver advice which is trusted and acted upon; and assessing the ability of satellites to measure small-holder yields within the context of a field experiment.

We designed an experiment to test the effect of site-specific fertilizer recommendations on fertilizer adoption and usage among cotton farmers in Gujarat, India, and evaluate our ability to measure yields. The sample for this experiment consists of 1,585 farmers who had recently signed up for a mobile phone-based agricultural advisory service, called Krishi Tarang (KT)[1]. The KT service provides comprehensive farming advice (on sowing, weeding, pesticides, harvesting, etc.) via automated voice messages throughout a cropping cycle. We randomly selected half of these farmers to receive plot-level soil fertility information and customized recommendations on three of the most common macronutrients and one micronutrient fertilizer (Zinc). These recommendations were delivered in a scalable manner: through the distribution of written information (Soil Health Cards and supplemental materials) prior to sowing, and a series of appropriately timed automated push calls during the growing season. To measure yields, we use both in-person surveys and satellite images. While remote sensing does require the mapping of individual plots, such mapping need not be done at the start of the experiment, and may potentially be useful for evaluating outcomes of interventions in settings where in-person surveying is difficult or expensive.

We examine the performance of satellites using two approaches to translate satellite data into yield predictions. The results speak to the need to exercise caution when using methods evaluated using metrics such as mean-squared error, which does not consider bias, in the context of evaluations, where bias is a first-order consideration. First, we follow a common practice of estimating the relationship between satellite vegetation indices and yields using OLS. The power gains to evaluating the intervention using this approach appear dramatic: we estimate that one could reduce the sample size needed to detect a 5% increase in yields by 73% if they switched from farmer-reported to satellite-measured outcomes using this approach. However, econometric theory suggests that this approach is likely to produce biased yield estimates in the presence of measurement error because attenuation bias will compress the distribution of satellite predictions. We propose the use of two-stage least squares (2SLS) to calibrate the yield prediction model as a solution to measurement error. Our results indicate that a researcher using OLS calibration would estimate a treatment effect $1/2$ of the true size. Furthermore, power calculations indicate that one would need a sample size 25% *larger* to detect a treatment effect using satellite data compared to farmer-reported data when using this calibration approach, if there is no differential attrition. However, satellite data may still offer cost savings by reducing field activities, and they offer large power gains if treatment effects are bounded to account for differential attrition.

With respect to the impact of site-specific fertilizer recommendations on farmer behavior change and yields, the

[1] KT service is provided by Precision Development (PxD).

evidence is more mixed. Both treatment and control farmers received general agricultural advice via mobile phone. The treatment group received additional advice on nutrient management, and most farmers listened to those messages. We find changes in self-reported fertilizer usage that are large within the context of the literature, but small relative to potential yield gains. Furthermore, neither noisy self-reported yields nor satellite data detect any yield gains from the intervention, perhaps in part because rainfall was particularly poor in our intervention year.

Returns to fertilizers are highly sensitive to dosage and heterogeneous by local conditions (Duflo et al., 2008; Suri, 2011), and governments in many countries are seeking to promote site-specific nutrient management. For example, since the launch of the Soil Health Card (SHC) scheme in 2015, the Indian government reports having conducted at least 50 million soil tests and delivered over 200 million SHCs with plot-specific fertilizer recommendations. Evidence suggests that Site-Specific Nutrient Management (SSNM) can lead to enhanced yields and improved soil quality under a controlled environment (Khurana et al., 2007; Pampolino et al., 2007; Cassman et al., 2002; Matson et al., 1998), though there is a paucity of literature examining whether locally-specific advice on fertilizer dosage and usage would lead to similar results in real-world, small-holder settings. ICT may be particularly well suited to improve the 'rate, timing, source and placement' of fertilizer, which could enhance the effectiveness of fertilizer application on plant absorption (Pagani et al., 2013).

This paper makes contributions to two strands of literature, (i) the use of remote sensing, particularly in experiments, and (ii) the promotion of agricultural technologies, such as fertilizer.

A recent body of research has demonstrated the potential of modern satellites to measure small-holder yields, and changes in yields, with greater accuracy than surveys. For instance, Lobell et al. (2020) compare full-plot crop-cut data, sub-plot crop cuts, satellite yield measurements, and farmer-reported yield data from small-holder maize plots in Uganda. They find that satellite yield measurements are about as accurate as sub-sample crop cuts when compared to full-plot crop cuts. For instance, satellite measurements explain 55% of the variation in full-plot crop cuts on pure stand plots larger than 0.1ha, whereas sub-plot crop cuts explain 47% of the variation in yields on these plots. Furthermore, satellite data is substantially more predictive of true productivity than farmer-reported data, and estimates of the returns to fertilizer are similar using satellite and crop cut measurements. Jain et al. (2019) show a high correlation between satellite wheat yield measurements and crop cuts in a split-plot experiment in India, and that satellite data can detect the effect of a fertilizer spreader in some instances, although point estimates are much smaller when measured with satellite data than crop cuts. Benami et al. (2021) provide an excellent overview of satellite measurements. They note that remote sensing measurements are less strongly correlated with ground measures in low-income compared to high-income economies, suggesting that satellites do measure yields with meaningful noise.

This paper examines the econometrics of satellite crop yield measurements in the context of treatment effect estimation. We show theoretically that the prevailing approach of calibrating predictions using OLS is not robust to measurement error, likely underestimating the magnitude of treatment effects. These predictions are consistent with the fact that work such as Jain et al. (2019) estimate smaller treatment effects using satellite data and the fact that the distribution of predicted yields is more compressed than in the case of crop cuts. We propose the use of instrumental variables in the calibration stage to produce unbiased yield measurements in the presence of measurement error in satellite or ground-truth data. In our setting, one would estimate treatment effects about $1/2$ as large when using OLS versus 2SLS calibration ($p < 0.01$). This calibration also has important consequences for statistical power calculations: we estimate that satellite data would reduce the sample size needed to detect a treatment effect by over 70% when using OLS calibration, compared to survey data, whereas 2SLS calibration leads one to conclude that a *larger* sample size is needed when using satellites. This contribution relates to Proctor et al. (2023), which documents that many remotely sensed variables in the United States produce biased parameter estimates. Our results similarly suggest that measurement error in satellite crop yield measurements can bias treatment effect estimates in experimental studies, but we propose the use of an instrumental variable approach which leverages knowledge about determinants of crop yields, rather than multiple imputation, to address this concern.

This paper builds on a large literature that seeks to evaluate agricultural extension. It is most closely related to Fishman et al. (2016), which evaluated in-person customized fertilizer advice in Bihar, India, but found that the advice did not affect the farmers' fertilizer application decisions. Fishman et al. (2016) propose limited trust and limited comprehension as barriers; our intervention seeks to overcome both factors. Harou et al. (2022) similarly evaluate plot-specific fertilizer recommendations, presented through in-person visits by agronomists, cross-randomized with fertilizer subsidies in Tanzania. The authors find that information alone has no effect on fertilizer use or yields, but combining soil fertility information with subsidies increases input use and yields relative to subsidies alone. Compared to Harou et al. (2022), we examine a setting where fertilizer use is common but imbalanced across types, and we examine a program that could readily scale to large numbers, using digital information rather than subsidies and in-person meetings.

We make three contributions related to agricultural extension. First, inspired by work on vaccines in Pakistan (Usman et al., 2011), we demonstrate that a human-centered design approach, emphasizing electronic delivery, generates more trust, and reported behavior changes, in a country where in-person visits have failed. In a "lab in the field" setting, Cole and Sharma (2017) find that the majority of cotton farmers in their survey sample in Gujarat had difficulty understanding recommendations in a government SHC but that digital and non-digital aid materials significantly improved comprehension levels. This paper uses an intervention very similar to Cole and Sharma (2017) and demonstrates that farmers understand, trust, and report acting on customized advice. To our knowledge, most literature on trust in digital information has focused on developed markets (e.g., Sillence et al. (2006), or Bonhard and Sasse (2006)).

More generally, the paper contributes to a growing literature exploring how advice delivered via mobile phones can increase technology adoption and improve agricultural outcomes (Casaburi et al., 2019; Fabregas et al., 2019; Cole and Fernando, 2020), and the literature on gains to fertilizer use more generally (e.g. Beaman et al. (2013)).

Finally, our study contributes to the literature on the possibility of using precision farming techniques in low-resource settings (Mondal and Basu, 2009; Maohua, 2001). On a positive note, we demonstrate that satellite data can yield estimates of the impact of interventions with similar precision to survey data, suggesting that this approach may help generate improved agronomic knowledge about the long-run effects of interventions, and that agronomic knowledge may be transmitted back to farmers at low cost. However, the results also serve as a cautionary tale.

Governments are increasingly investing resources to build databases on local conditions and farmer characteristics, with the goal of providing increasingly customized advice. Our findings suggest that further research is warranted to verify that the investment in soil health cards and soil mapping would bear fruit: while we find that farmers report changing behavior, we do not observe yield gains.

2 Background and Experimental Design

While fertilizer use is widespread in India (over 75% of cultivated land uses fertilizer[2]), farmers have limited access to advice about optimal quantities. For example, only 6% of farmers reported interacting with an extension agent in the previous year (Cole and Sharma, 2017), and respondents in our baseline survey answered only 2.3 of six questions about fertilizer correctly.

To improve soil health management, the Government of India launched a Soil Health Card scheme in 2015, with the goal of testing soil at fine spatial resolution[3], and delivering personalized SHC to every farmer in India (`https://soilhealth.dac.gov.in/`). However, Cole and Sharma (2017) note potential shortcomings with the government's approach. Not all farmers may receive soil health cards. Moreover, the technical presentation of data is difficult to understand, and only 8% of farmers in their sample in Gujarat understood the basic recommendations on the SHC.

2.1 Intervention

This study was implemented in partnership with Precision Development (PxD), an NGO providing mobile phone-based agricultural extension services to small-holder farmers. PxD operates Krishi Tarang (KT), a two-way voice-based advisory service. Farmers subscribed to the KT service receive weekly push calls with information on seeds, pesticides, planting, harvesting, and other agricultural decisions. Farmers can also call back into the system to access their personal inbox, re-listen to the messages sent through push calls, and record any questions, which would be answered by a PxD agronomist within two days. The service is offered for free and currently has more than sixty-thousand active users, overwhelmingly growing cotton, across thirty-six districts of Gujarat, a state in western India.

Cole and Fernando (2020) evaluate an early version of the service, finding high demand for the service and evidence of systematic changes in agricultural practices. An important potential advantage of ICT delivery advice is the ability to deliver individually customized advice. To examine the feasibility of such an approach, PxD designed a set of messages that explain the importance of soil fertility management and provided farmers with information on plot-level soil nutrient levels, benefits, and recommended dosages of three macronutrient (UREA, DAP, MOP) and one micronutrient (Zinc) fertilizers. The farmers are not tightly integrated into a value chain. But the nutrients we recommend were readily available in our study area, in sufficiently small quantities that indivisibility constraints would not bind.[4]

[2] Input Survey, 2011
[3] 2 x 2 hectare grid for irrigated farmland
[4] Only UREA was recommended for unirrigated cotton.

The specific intervention we study is inspired by government campaigns to inform farmers about their soil health, but designed to offer additional supportive content. For all farmers in our study, PxD sent field staff to collect a soil sample from the farmer's primary cotton plot as part of the baseline survey[5]; soil tests were performed by a local agricultural university; and PxD applied the university's fertilizer model, which is calibrated to maximize profits, in combination with soil test results to generate fertilizer recommendations[6] customized for the individual plot. Recommendations depend on irrigation status, which can vary over the course of the season in this sample, because many irrigation sources depend on rainfall. We use reported irrigation status at the time of basal fertilizer application to determine whether farmers followed basal recommendations, and midline irrigation status to determine whether they followed full season recommendations.

Treated farmers received customized recommendations through multiple channels while control farmers did not receive soil health results during the study period. At the start of the agricultural season, PxD hand-delivered a Soil Health Card (SHC) and two supplementary, customized, printed materials to each treated farmer in our sample. To ensure that recommendations were understandable and actionable, we simplified the design of SHC using an iterative process of testing the comprehension level and tweaking the design based on feedback from farmers in the study area while maintaining the amount of information provided in the SHC (Appendix Figure 1).

In addition, supplemental materials were designed to help farmers understand the fertilizer recommendations: a card (Appendix Figure 2) that lays out the timing and the quantities of different fertilizers recommended without detailed information on nutrient values and a booklet (Appendix Figure 3) that provides pictorial illustrations of the potential effects of each fertilizer type on plant health and yields.[7]

Finally, appropriately timed recommendations on fertilizer application were delivered through push voice calls over three months between June and September, incorporated into PxD's advisory service. Push calls were timed to coincide with key stages in the crop growth cycle. In each call, PxD announced the topic of the call (macronutrient fertilizers, MOP, or zinc fertilizers), specified whether the call was for irrigated or unirrigated cotton, explained the potential benefits of the fertilizer(s), and provided the recommended application quantities (recommendations were given in amount per area unit). If farmers did not pick up the call on a scheduled day, calls were sent again the day after. On both days PxD made three attempts to reach farmers if they did not pick up the call. The share of farmers that listened to the push call recommendations, presented in Appendix Table 5, was greater than 70% for every topic except for mid-season zinc application.

2.2 Sample frame

The sample of this study consists of cotton farmers who registered into the KT service in the first quarter of 2018 across three districts (Surendranagar, Rajkot and Morbi) in the southwestern region of Gujarat. PxD administered a

[5]Farmers were informed that they would receive the SHCs within 10 months. The research team visited farmers in the control group after the experiment to distribute SHCs.

[6]While in theory PxD could have drawn information from mass soil tests conducted by the government, previous work (Cole and Sharma, 2017) raised concerns about the accuracy of government soil tests. Technical details of the process of generating fertilizer recommendations are summarized in Appendix Table 1.

[7]As the spread of smartphones and data plans increases, we imagine these visual aids could be delivered in video format at low cost.

screening survey and identified farmers who owned a mobile phone, were planning to grow cotton in the upcoming Kharif season, and were interested in receiving agricultural information through the KT service but had not subscribed to the service before. The baseline sample, including all farmers who had a plot suitable for soil collection, was 1,585 farmers.[8]

2.3 Randomization and sample characteristics

Farmers in the base sample were stratified by block (district subdivision) and randomly assigned to a treatment or control group (793 and 792 respectively). Table 1 presents summary statistics of the key variables at baseline by experimental groups. Column (1) reports the control group mean and standard deviation of these variables. The average age of farmers in the study was 43 years, over 95% of the sample indicated that their primary occupation was self-employed farming, and more than eighty-percent of respondents were literate (could read a newspaper in the local language). The average farmer had a total cultivated land size of about 3.5 hectares the year prior to the intervention, slightly above the national average of 2.35 hectares among all cotton farmers in India. Even though the Indian government launched a nationwide Soil Health Card scheme in 2015 with the plan of conducting a soil test for every farmer, fewer than 15% of farmers in our sample reported ever having their soil tested. The remainder of the table reports balance checks, for the randomization (Column 2), and for comparisons based on baseline characteristics among the sets of farmers who did not attrit in subsequent surveys: a basal survey (conducted a few days after PxD recommended applying the first application of fertilizer), a midline survey (conducted in October-December 2018), an endline survey (conducted in February-May 2019), and a separate plot mapping exercise (conducted in March-May 2019). The p-value of an F-test of the joint orthogonality of the variables is included at the bottom of Columns (2) - (7).

Table 1 shows that the proportion of baseline variables imbalanced between the two experimental groups is below the corresponding significance levels, indicating that the two groups are well-balanced. We will discuss differential rates of attrition in Section 8. In addition to the survey data, we obtain administrative data on KT service usage including pickup rates and amount of time spent listening to messages. These data are available for both treatment and control groups, though only treatment groups received supplemental fertilizer advice.

2.4 Empirical strategy

We estimate an ITT with OLS:

$$Y_i = \alpha_b + \beta T_i + X_i'\theta + \epsilon_i \tag{1}$$

where Y_i denotes the post-intervention outcome for individual i, T_i the treatment indicator, b the sub-district fixed effects, and X_i is a vector of controls. We typically do not include controls when estimating treatment effects on fertilizer use, whereas we do control for lagged productivity measurements when examining effects on yields to improve precision.

[8] A soil sample could not be collected if there were standing crops from the previous season or if fertilizer had been applied after harvesting of the previous year's crop.

One goal of this paper is to better understand whether and how point estimates and standard errors change depending on the methodology used to measure outcomes. To the extent possible, we seek to separately identify differences due to changes in sample composition vs. changes in the noisiness of the measure. To that end, we report the results of several regressions across four samples. First, we present results across the full sample. Second, we present results across the sample of respondents that provided yield data. This sample size is equal to 1,341, which is 61 observations smaller than the set of respondents that completed the endline survey because some farmers did not provide productivity data during the survey. Third, we restrict the sample to the 1,326 respondents for which we have 2018 satellite yield data. Fourth, we examine the set of 1,291 respondents for which farmer-reported and satellite yield data are available.

3 Results

3.1 Take-Up: Listening Rates of Customized Fertilizer calls

Take-up of the information service, detailed in Appendix Table 5, was quite high, with 76% (unirrigated) to 88% (irrigated) of farmers picking up and listening to basal fertilizer recommendations. In fact, 50% of the treated farmers listened to the same recommendation more than once. Pick-up rates were lower for calls later in the season, but the listening rates for all but one of the recommendations remained above 70%. In total the average treatment farmer was exposed to 42 minutes of content on fertilizer application.

However, we observe some "crowd out" of information, as treated farmers are 3.5 percentage points less likely to pick up subsequent weekly calls on cotton farming advice (Table 2); this difference is significant at the 1% level. However, Column (2) shows that, unconditional on picking up, treatment farmers listened to approximately the same amount of total non-fertilizer content, meaning that the total exposure to content was roughly equivalent.

The intervention appears to have increased trust in the KT service. Treated farmers rated non-fertilizer KT calls higher by an average of 0.148 points on a 5-point scale, and they were 3.3 percentage points more likely to report using a mobile phone-based advisory service to make agricultural decisions during the endline survey. The intervention also increased reported trust in mobile phone-based advisory by over 0.2 points on a 5-point scale ($p < .01$).

3.2 Impact on fertilizer knowledge

Surveyors administered a 14-question multiple choice quiz about basic facts relating to fertilizer during the midline survey to measure whether the treatment increased farmers' fertilizer knowledge (See Appendix Table 2 for questions). Baseline knowledge was low, with farmers correctly answering approximately 3.2 questions on average. The intervention increased fertilizer knowledge: farmers in the treatment group provided the correct response to 0.266 - 0.316 more questions, depending on the sample frame. The knowledge gains are statistically significant at the 1% level across all samples. Taken together, the high adoption of the customized fertilizer recommendations, increased trust in mobile phone-based advisory, and improved fertilizer knowledge indicate that trust and comprehension of the fertilizer recommendations were high. The intervention was thus able to alleviate some of the information frictions that Fishman et al. (2016) identified as barriers to the adoption of government SHC recommendations.

3.3 Impact on fertilizer use

We next examine whether the customized fertilizer recommendations improved fertilizer use. The treatment push calls heavily emphasized fertilizer application at the basal (first) dose, which agronomists at PxD thought may be particularly inefficient. As a result, treatment effects on fertilizer use are reported for the basal dose and in aggregate.

In Table 3, we evaluate the effect of the experiment on self-reported basal fertilizer use, using three metrics: a binary measure equal to one if the farmer reported using fertilizer that was recommended, or did not report using a fertilizer that was not recommended, and zero otherwise; the total quantity of fertilizer applied; and the absolute value of the difference between the amount recommended and the amount used. To simplify analysis and address multiple hypotheses testing, we calculate a standardized index of fertilizer application, which is the equal-weighted mean of standardized effect sizes for the four inputs (UREA, MOP, DAP, and Zinc). Column (1) indicates that treated farmers were 0.232 standard deviations ($p < .01$) more likely to follow recommendations about which fertilizer types to apply across the full sample. This effect was driven by UREA and MOP (Table 5). Column (2) shows that treated farmers applied more fertilizer on average, with standardized joint effects of 0.363 standard deviations or larger across each sample. Each difference is statistically significant at the 1% level. In terms of quantities, average usage increased by 11.4 kg/ha for UREA, 9.6 kg/ha for MOP, and 0.8 kg/ha for Zinc, corresponding to increases of 140%, 435%, and 1,061% over average doses in the control group. This was primarily due to changes at the extensive margin. Basal application of DAP decreased by 4.7 kg/ha, but the value is not significantly different from zero. Column (3) shows the intervention did not promote overuse of fertilizer: there was a statistically significant decrease in the distance between the recommended fertilizer dose and the reported fertilizer dose, which dropped by 0.122 standard deviations averaged across the four fertilizer types. Table 5 and Figure 2 break down changes in the fertilizer gap by fertilizer type. This table is intended to provide evidence about which fertilizers drove changes observed in the standardized joint effects, and that readers should be cautious interpreting changes in individual fertilizer usage absent a difference in the joint effects due to multiple hypothesis testing. The largest reduction in the fertilizer gap occurred in the case of UREA, driven by a decrease in under-application.

Table 4 shows that there are similar changes for full-season fertilizer application by treated farmers. Total fertilizer application increased by an average of 0.237 standard deviations across UREA, DAP, MOP, and Zinc among the full sample ($p < .01$). The standardized joint effects are all 0.239 standard deviations or larger among the three restricted samples. Moreover, there is a statistically significant decrease in the standardized joint effect of the difference between the suggested and applied fertilizer quantities of at least 0.076 across each of the four samples. We estimate that farmers in the treatment group spent an average of about INR 300 more on fertilizer over the course of the season, relative to a base of about INR 8,300 in the control group. However, differences in fertilizer expenditures are not significantly different from zero. We suspect that this may be attributed in part to the fact that fertilizer expenditure data is noisy, but the lack of a significant effect is also due in part to weighting of the fertilizer types. We are able to detect an increase in the standardized joint effect on fertilizer usage because this statistic weights MOP usage, which had a large treatment effect and a small standard deviation due to low control usage, highly, whereas DAP and Zinc had the highest prices so factor most strongly into revenue calculations, and there was not a statistically significant increase in the amount of these fertilizers applied.

Although the intervention led to improved fertilizer application overall, there is little evidence that macro- and micro-nutrient fertilizer use was more balanced among treated farmers. Table 5 and Figure 3 indicate that increases in the use of UREA (a 15% increase) and MOP (a 306% increase) drove changes in total fertilizer application. There is not a statistically significant change in Zinc application measured in kg/hectare, although treated farmers were almost twice as likely to report applying any zinc ($p < .01$).[9] In contrast to reports of UREA overuse, Figure 3 suggests that the majority of our study sample under-applied UREA. The Government of India's SHC dashboard also indicated that over 80% of government soil tests found that soil nitrogen levels were low or very low as of 2020, suggesting that optimizing UREA application could improve farm productivity.[10]

We do note, however, that fertilizer use may have been lower in the control group during the study year because of low rainfall. Appendix Table 6 plots extensive and intensive margin fertilizer use for the study year. We observe large control group declines in the use of UREA and MOP from 2017 to 2018, the two most commonly applied fertilizers.

Another limitation of our analysis is that we do not observe multiple years of fertilizer use. The fact that we observe large increases in zinc adoption at the extensive margin but not in overall application amount implies that treated farmers that applied zinc used less on average, although this difference is not statistically significant. One possible explanation is that the intervention increased experimentation with micro-nutrient fertilizers that farmers were not familiar with.[11] Hence, one might expect treatment effects to change if the intervention were continued, although this is only speculative since we lack the data to rigorously test this hypothesis.

3.4 Treatment effect on yields

We next examine whether improvements in fertilizer application across the basal dose and full season, driven by reductions in the under-application of fertilizer, led to an increase in cotton yields.[12] Table 4 shows that the point estimate of the effect of the intervention on farmer-reported yields is just under 1 kg/ha across the full sample, a productivity change of about 0.1% that is not statistically significant, no matter the specific sample restrictions we use. Figure 4 similarly shows no difference in yields between the treatment arms.[13]

Our primary focus is on yields, not profits, because we collected more detailed data about yields. However, Appendix Table 7 reports treatment effect estimates on cotton sale revenue which suggest that there was not a positive treatment effect on cotton sales revenue, consistent with the lack of a treatment effect on yields. In fact, the treatment effect on sales revenue in column (3) and expected sales revenue – which includes estimates of revenue from stored

[9] The extensive margin change in Panel D of Table 5 is only marginally significant since it equals 1 if farmers did apply zinc and had irrigation at midline or did not apply zinc or have irrigation at midline. The share of control farmers that applied zinc is 5%, and the treatment effect is 4%. Farmers that applied zinc without irrigation generally expected to have irrigation, but their system was dependent on rainfall so their crop was unirrigated at midline.

[10] `https://soilhealth.dac.gov.in/NewHomePage/StateWiseNPKChart`

[11] All farmers that applied zinc reported applying it to their full plot, so we are not detecting experimentation on a subset of the plot.

[12] Strictly positive yield values were winsorized at the 2nd and 98th percentile to reduce the influence of outliers at the low and high ends of the distribution.

[13] An agronomist at PxD estimated that observed fertilizer differences could translate into 3-4% yield gains which we are not powered to detect.

cotton – in column (4) are negative and statistically significant, but these effects would not survive corrections for multiple hypothesis testing. The estimated effect on revenue net of fertilizer costs (column 6) is positive but not significantly different from zero. We did not collect detailed data about inputs unrelated to fertilizer use. We thus view the results in Appendix Table 7 as suggestive evidence that the intervention did not increase profits, but cannot reject the null hypothesis that the treatment had a positive effect.

Although the point estimate of the effect of the intervention on yields is close to zero, using farmer-reported yields results in a relatively wide 95 percent confidence interval, ranging from -77.95 kg/ha (-8.2%) to 79.93 kg/ha (8.4%). As a result, we cannot reject a large increase in agricultural productivity, which, if true, could admit positive benefit-cost analysis. In the next section, we evaluate whether satellites can estimate more precise bounds on the treatment effect.

4 Using satellites to estimate treatment effects

In this section, we use satellite imagery to measure agricultural productivity and examine the performance of remote sensing data relative to farmer-reported outcomes. Detailed information about the data and methodology used to construct satellite yield estimates is presented in Appendix I. In addition, we include a technical appendix, which contains a "cookbook" with step-by-step instructions for constructing satellite yield measurements; the instructions are designed to be accessible to those without prior remote sensing knowledge, coding experience, or specialized computing resources. Appendix Figure 4 outlines the steps required to calculate yield measurements using these tools. The technical appendix includes commented code, which is also accessible in a GitHub repository at `https://github.com/gkilleen33/rs-economics`.

Briefly, satellite yield estimates were constructed from Sentinel-2 imagery, which spans the globe and is available at no cost from the European Space Agency. Yield measurements were constructed as follows. First, we constructed five vegetation indices (VIs) from the multispectral imagery: NDVI, GCVI, reNDVI, MTCI, and LAI.[14] Second, we calculated the median value of each VI for each plot (specifically, over the portion of the plot containing cotton) and each satellite image. Third, we took the maximum VI value observed in each plot across a growing season. This produced a data set containing one measurement of each VI per field in the 2016, 2017, and 2018 growing seasons. We use the maximum VI value, rather than a different statistic of the VI values, because the maximum is more robust to differences in sowing time.[15] We then evaluated the performance of each VI individually by comparing VI measurements to farmer-reported yield data. Finally, we selected the best-performing index, and we constructed yield estimates by linearly fitting the VI values to farmer-reported productivity. As discussed later in this section, a common practice is to calibrate the linear model using an OLS regression. However, we present evidence that this approach produces systematic bias in yield predictions which is likely to attenuate treatment effect estimates. As a result, we

[14]Leaf Area Index (LAI), which is calculated using a neural network, is not technically a vegetation index. We describe it as a VI since it provides a similar measurement of productivity.

[15]The maximum VI over the course of an agricultural season has been found to be highly predictive of yield in the remote sensing literature (e.g. Lambert et al. (2018)). An exception might be if a large proportion of healthy crops is destroyed prior to harvest, for example, due to an extreme weather event. This did not occur in our study.

propose the use of 2SLS to estimate the calibration model.

4.1 Satellite yield measurement

We begin by evaluating whether satellite imagery can produce accurate estimates of cotton productivity in this sample. Table 6 demonstrates that there is a strong correlation between farmer-reported productivity and the satellite VI measurements. The table presents regressions of the five VIs that we calculated on farmer-reported 2018 yield. The coefficients on yield are all positive and statistically significant, and the R^2 exceeds 0.2 for each model and peaks at 0.287 in the case of reNDVI. Figure 6 confirms a strong and positive relationship between the VI values and farmer-reported yield.

It is important to ensure we are able to pick up plot-level variation (rather than regional yield variation). First, we note that including very fine geographic fixed effects (0.5 km x 0.5 km blocks) only marginally affects the slope, and does not affect the statistical significance level of the relationship between the VIs and reported yields (the p-value on the vegetation index with 0.5km x 0.5km fixed effects is reported below the "Results with grid FE" header). Second, we perform a placebo test in which we calculate satellite VI values using dummy plot boundaries that contain areas close to each field in the sample, but do not contain areas from the sample plots. Specifically, each plot boundary polygon is replaced with a torus containing the area 100 meters from the plot to 200 meters from the plot. The relationship between the placebo VI values and farmer-reported yields is still positive and statistically significant, reflecting spatial differences in average productivity, but the R^2 declines substantially relative to the VI values calculated using the actual plot boundary data. These values are reported at the bottom of Table 6. For instance, the R^2 between reNDVI and farmer-reported yield using the plot boundary data is 0.287, but it is only 0.134 when the placebo data is used. When we use the placebo plots and include 0.5 x 0.5 km grid fixed effects, there is no statistically significant relationship between yield and any of the VI values. These results are consistent with the ability of satellite data to differentiate between the productivity of individual plots in this sample.

In the technical appendix, we further demonstrate that the fit between farmer-reported yields and satellite VIs is better on large plots than small plots (Appendix Table 10) and that satellites detect the increased importance of irrigation during the 2018 drought (Appendix Table 11). Overall, these findings demonstrate that satellites can provide a reliable measure of cotton productivity in this sample. The best-performing VI is reNDVI, so we calculate satellite yield estimates using this index.

In order to convert from VI units – which lack an economic interpretation – to yields, one needs to calibrate a prediction model. Based on prior research observing a linear relationship between NDVI and yields, such as Stamatiadis et al. (2010), and the fact that the relationship between reNDVI and yields appears to be linear in Figure 6, we model this relationship linearly

$$Yield_i = a_0 + b_0 \cdot reNDVI_i + \nu_i \tag{2}$$

where $Yield_i$ is farmer-reported yield in kilograms normalized by GPS-measured plot size in hectares. A common approach in prior work has been to estimate this calibration model via OLS (Burke and Lobell, 2017; Lambert et al., 2018; Lobell et al., 2019; Jain et al., 2019; Lobell et al., 2020, e.g.). Indeed, an earlier version of this paper took exactly this approach; the journal's referees raised concerns that this approach may understate variation in satellite yield measurements. This prompted us to reconsider the use of OLS calibration; we argue that $\hat{b}_{OLS}$ is likely to be biased.

There are two sources of potential bias from OLS. First, if the true structural equation is in fact

$$reNDVI_i = a_0 + b_0 \cdot Yield_i + \nu_i \tag{3}$$

then applying OLS to Equation 2, we would estimate $\hat{b}_{ols} = \frac{1}{\mathbb{E}[\nu_i^2]} b_0$, a standard result that estimating reverse regressions via OLS produces bias. Equation 3 may be interpreted as stating that the observed vegetation index is a function of yields and noise, an interpretation that we view as reasonable. Second, measurement error in reNDVI would produce attenuation bias in $\hat{\beta}_{OLS}$. This is likely since factors such as atmospheric haze are important in yield predictions and satellite pixels are large, so they often overlap with areas outside of the plot.

A biased estimate of b_0 does not necessarily produce imprecise estimates of yield. In fact, James and Stein (1992) show that using estimators which are biased towards zero can reduce mean-squared error due to a bias-variance trade-off. But if we consider the treatment effect on reNDVI,

$$reNDVI_i = \gamma_0 + \gamma_1 T_i + \omega_i \tag{4}$$

then one can easily show that estimating the treatment effect of the intervention on predicted yields results in the estimate

$$\hat{\beta} = \hat{b} \cdot \hat{\gamma}_1 \tag{5}$$

By randomization, we know that $\hat{\gamma}_1$ is an unbiased and consistent estimate of γ_1. But if our estimate of b_0 is biased, for instance in the case of $\hat{b}_{ols}$ then the estimated treatment effect on yields will be biased. Intuitively, if one underestimates the elasticity of yields with respect to reNDVI, then treatment effects on yields will be underestimated because the vegetation index is not scaled accurately to the units of economic interest. In the context of estimating economic parameters, producing an unbiased estimate of b_0 is of first-order importance.

Proctor et al. (2023) introduces a multiple imputation methodology to correct bias in a broad set of remotely sensed outcomes in the United States. While in principle this approach could work when predicting smallholder yields, it requires a relatively large volume of accurate ground truth data, which may be scarce or very expensive to obtain in developing country settings. Instead, we propose the use of 2SLS to estimate the calibration stage in agricultural settings. Assuming valid instruments for the VI, this produces a consistent estimate of b_0 and so estimation on predicted yields will be consistent. Compared to the multiple imputation approach, this technique may be more robust to measurement error in yields, requires fewer observations, and is conceptually easier to interpret. Furthermore, Lewis and Linzer (2005) demonstrate that as long as predicted yields are unbiased, then researchers need not explicitly account for uncertainty in the calibration stage when calculating standard errors about treatment effects so long as heteroskedastic-robust standard errors are estimated. Hence, the IV approach produces no loss in statistical power compared to OLS calibration. We show that total rainfall over the growing season, observed from satellites, and sowing date are strong instruments for reNDVI and that they produce a statistically and economically larger estimate of b_0, consistent with substantial bias from the use of OLS calibration.

Appendix I reviews the emerging literature on satellite yield measurement. Relative to existing literature, this paper makes the following contributions:

First, we consider econometric challenges unique to the use of satellites in treatment effect estimation, particularly in RCTs. Much of the existing work (eg Burke and Lobell, 2017; Lambert et al., 2018; Lobell et al., 2019, 2020) evaluating the potential of satellites to measure crop yields has focused on statistics such as R^2 or root mean-squared error which are informative about the ability of satellites to detect a signal about crop yields. However, James and Stein (1992) demonstrate that due to the bias-variance trade-off, estimates that minimize prediction error need not be unbiased. We present evidence that the common practice of calibrating prediction models with OLS is likely to produce substantial bias in treatment effect estimates, and we show that calibrating yield measurement using a set of instruments that is widely available in agricultural studies can produce consistent estimates. These findings can rationalize the fact that Jain et al. (2019) find smaller treatment effects and a compressed distribution of yields with satellite data.

Second, we explore the effects of replacing farmer-reported yield data with satellite imagery on statistical power using both OLS and 2SLS calibration. We find that, in the absence of differential attrition, one would be able to detect a 5% yield increase with about 30% of the sample using satellite data calibrated with OLS compared to farmer-reported outcomes. However, once we use 2SLS calibration that we argue is unbiased, the story flips and satellite data has worse power to detect a treatment effect. Nonetheless, satellites may still offer advantages. Given corrections for differential attrition, of which we find little evidence in this sample, satellites offer large power gains since survey non-response is impossible. Even without differential attrition, a 30% increase in sample size to use satellite data may result in overall cost-saving by allowing for fewer rounds of on-ground data collection. In addition, the use of satellites also allows for long-run monitoring with no additional fieldwork.

Third, we provide a "cookbook" to assist researchers without remote sensing expertise in generating satellite yield estimates and applying them to RCTs. We aim to make satellite yield measurement methods evaluated in previous studies accessible to other disciplines so that the benefits of these advancements can be more widely realized. The "cookbook" consists of detailed instructions and code presented in the technical appendix as well as a code repository that is available at `https://github.com/gkilleen33/rs-economics`. We aim to update the repository to reflect the best available methods, and we welcome other authors to contribute to it. None of the resources presented require remote sensing or coding knowledge, and they use computing resources that are free to academics when possible.

4.2 OLS vs 2SLS yield calibration

We first examine how yield predictions vary if we calibrate the model using OLS versus 2SLS in Table 7 and Figure 7. All estimates include block fixed-effects. We exclude observations for which farmer-reported yield is 0, typically meaning their crop failed, or greater than or equal to 3,900 kg/ha since these represent large outliers that may reduce our ability to estimate the relationship. Two-stage least squares estimates use total rainfall from June to October and sowing date – variables which are easily obtained in most agricultural settings – as instruments. These instruments will produce a consistent estimate of the relationship between reNDVI and yields if they are correlated with yields only through their relationship with reNDVI. Since reNDVI is a measure of biomass, we believe this assumption is credible: the instruments are likely correlated with yields via their effects on vegetation but uncorrelated with sources of measurement error like atmospheric interference for specific satellite images. Sowing date could be correlated with measurement error if it were picking up the fact that farms were imaged by satellites at different stages of the growing cycle. In our case, we believe this is unlikely since we take the maximum reNDVI value across passes to integrate out differences in reNDVI due to different crop stages. Hence, we expect sowing date to be correlated with the maximum of the index primarily due to its effects on the timing of the plant's exposure to rain, affecting biomass and yields.

Columns (1) and (3) of Table 7 report direct estimates of the relationship. Using OLS, we estimate that a 1 unit increase in reNDVI corresponds to a yield increase of about 4,000 kg/ha. In contrast, the 2SLS estimate is over 10,000 kg/ha, a difference between coefficients that is significant at the 1% level. The selected instruments are strong, the first stage F-statistic is over 60, and a J-test of over-identifying restrictions fails to reject the exogeneity of the instruments. We interpret this as evidence that the instruments are valid, indicating that researchers that use OLS calibration are likely to estimate treatment effects under half of their true size. We note that since weak instruments or violation of the exclusion restriction would likely bias the relationship towards OLS, one would likely still conclude that OLS calibration is biased even if they were not convinced by our instruments.

We further test whether measurement error binds by estimating reverse regressions, and inverting the estimate so they are comparable. These results are reported in columns (2) and (4). We find that the 2SLS estimates do not meaningfully change, but the OLS coefficient more than doubles and is similar to the 2SLS estimates when we estimate the reverse regression and then invert it. This supports the interpretation that reNDVI is equal to yields plus noise, and that a calibration via linear regressions of yields on reNDVI would be biased.

We next plot predicted yields versus farmer-reported yields in Figure 7. Panel A plots predictions based on OLS calibration. The distribution of predicted yields appears to be compressed compared to the farmer-reported estimates, overestimating low-yield values and underestimating high-yield values. This is consistent with attenuation bias in the calibration step. In contrast, the predictions based on 2SLS in Panel B appear to fall along the 45-degree line, consistent with an unbiased estimate. Furthermore, the maximum of yield predictions is similar to the maximum of farmer-reported values, although some negative yield values are predicted.

4.3 Treatment effect on yields: Satellite vs farmer-reported data

We next examine how treatment effect estimates and standard error vary with farmer-reported versus satellite yield data in Table 8 and Figure 8. The sample is restricted to observations for which none of the data sources is missing so that we can directly examine how noise in the data affects estimates. The role of attrition is considered separately in the following section.

Column (1) of Table 8 reports estimates in which the outcome is farmer-reported yield normalized by farmer-reported plot size. Column (2) normalizes farmer-reported yields by GPS-measured plot size. Column (3) examines satellite yield data calibrated with OLS, and column (4) uses satellite measurements calibrated with 2SLS. Each of the regressions includes block fixed-effects and controls for lagged productivity measurements.

We find that neither point estimates nor standard errors of treatment effects change substantially when we normalize farmer-reported yields by farmer-reported versus GPS-measured plot size.[16] Standard errors do change substantially if we switch to satellite-measured outcomes, however, and the direction of change varies depending on whether we use OLS or 2SLS calibration. In column (3), we see that standard errors drop by about 50% when using OLS-calibrated satellite measurements. However, standard errors are over 20% larger in column (4) when considering measurements calibrated with 2SLS. This is consistent with the fact that OLS calibration compresses the distribution of yields, leading researchers to produce point and standard error estimates of treatment effects that are biased towards zero. Figure 8 presents similar findings visually.[17]

4.4 Power calculations: Sample size changes with satellite imagery

In this section, we present power calculations that estimate how the use of satellite data would affect statistical power in a setting in which plot boundary data were collected prior to the beginning of the intervention. We perform power calculations using both OLS and 2SLS calibrated yield predictions to examine how the calibration decision affects trade-offs between data sources. Unlike in the previous section, we carefully consider attrition. We further examine how different sets of controls and differential attrition enter considerations.

Table 9 presents the sample size needed to detect a 5% change in cotton yields 90% of the time with 95% confidence. Each sample size is estimated using a bootstrapping procedure described in detail in Appendix 8. In brief, we draw 1,000 bootstrapped samples on a grid, then calculate the share of times that a test rejects if we impose the alternative hypothesis. Column (1) reports estimates using farmer-reported data normalized by farmer-reported plot size. Column (2) examines satellite predictions with OLS calibration, and columns (3) - (4) examine satellite data with 2SLS calibration. Column (4) differs from column (3) in that for each bootstrap draw the calibration is re-estimated, whereas columns (2) and (3) use the calibration measured on the full sample. In practice, we find that this decision does not significantly affect power calculations.

[16] Farmer-reported and GPS-measured plot sizes are strongly correlated in this sample. The R^2 between the measures is about 0.8, the slope is approximately 1, and the intercept is not different from 0 by a statistically significant margin.

[17] We also tested for treatment effects on satellite-measured yield the following season and found no treatment effect. We omit this analysis because we did not collect data on fertilizer use the following season, and the control group received soil health cards at the end of the study year, so we cannot determine whether we fail to detect an effect because there were no differences in fertilizer use or because there were differences in fertilizer use but they did not increase yields.

The first row of the table presents sample sizes assuming random attrition controlling for all available lags of the outcome. We estimate that one would need a sample size of over 12,000 to detect an effect with farmer-reported data, a well-documented fact that large sample sizes are needed to study agricultural effects because yield data is noisy. Column (2) indicates that one could detect an effect with a sample size of about 3,500, about 70% smaller, if the OLS calibration were unbiased. However, columns (3) - (4) raise the concern that these power gains may be due to biased calibration compressing the distribution of yields: the sample size needed to detect an effect *increases* to over 15,500 if we consider satellite data calibrated with 2SLS.

Row 2 of column (4) examines how power changes if we only control for 1 lag of satellite data. We find that the sample size increases to over 19,000, demonstrating the value of ANCOVA regressions (McKenzie, 2012). We omit this calculation for the other two satellite measures due to computational power. If we control for no lags of the dependent variable, the sample size in column (4) increases to about 21,000, and the sample size with farmer-reported data increases to about 12,647. Hence, lagged controls appear to be more valuable with satellite than farmer-reported data.

The estimates in rows 1 - 3 assume the attrition is random. Appendix Table 4 reports the likelihood of attrition due to non-response to the basal survey (Column 1), midline survey (Column 2), endline survey (Column 3), farmer-reported yield questions (Column 4), and plot mapping exercise (Column 5) conditional on treatment status, baseline characteristics, and interactions between treatment and the baseline variables. The coefficients on treatment are all statistically insignificant, suggesting there is no differential attrition. In addition, the interaction terms between treatment and the baseline variables are not jointly different from zero by a significant margin, with the exception of the basal survey ($p = 0.058$).

However, in the real world, studies which deliver interventions are often subject to differential attrition. We therefore examine how corrections for even the small levels of differential attrition due to survey non-response, as we observed in our sample, affect power calculations, in the final row of Table 9. (Some cells in this table are blank as non-random attrition is not realistic in satellite data). We estimate the lower Lee (2009) bound for each bootstrap, and then determine whether the test still rejects. We estimate that one would need a sample size of over 34,000 if they apply corrections for differential attrition and still have 90% power. Sample size falls dramatically because yield data is noisy, so trimming has large effects on estimates. Hence, in the presence of differential attrition satellite data may offer large benefits. However, our results indicate that with random attrition and 2SLS calibration, satellites do not more precisely estimate treatment effects than farmer-reported data in this sample.

There are, however, several additional reasons that satellites may perform better relative to survey data in other settings. For food crops, household consumption may introduce measurement error; this is not a concern in our setting, with cotton. Similarly, farmers observe the weight of their yield when selling the crop, which may lead to more precise survey data on yield. Second, differential attrition can pose serious threats to the validity of impact estimates using survey data for a range of agricultural interventions and in settings where temporary migration is common. Third, only two years of pre-intervention Sentinel-2 data existed at the time of this study, because of the date that the satellite constellation was launched, but future work would benefit from more years of pre-intervention data, leading to improved precision and reduced required sample sizes. Fourth, given the natural, large variation in agricultural outcomes, multiple seasons or years of post-intervention outcome data can substantially improve our ability to understand the impact of agricultural interventions. Fifth, satellite measurements may become more accurate as spacecrafts and the software used to process satellite imagery improve. Hence, we view our results as evidence that satellite yield measurement may not outperform survey data in some settings. But our results still present promising evidence that current publicly-accessible imagery sources can detect yield differences across smallholder plots, and are likely able to detect treatment effects with sample sizes only modestly larger than survey data in this setting. Depending on the characteristics of the target crop and the risk of differential attrition, satellites may be a more cost-effective substitute for household surveys, or they may be a useful complement to in-person surveys.

In addition, satellite measures can be useful even when survey measures are used. For instance, satellites could be used to predict what crops were sowed on the plots of farmers that did not participate in follow-up surveys. They can also test whether yields are lower on these plots. These data points could help test for the effects of differential attrition, avoiding the need to bound estimates.[18]

4.5 Costs of satellite yield measurements

Despite the fact that satellite data requires a sample size about 30% larger to detect a treatment effect, the data source may still offer cost savings by reducing field activities. We estimate that plot mapping increased the per-farmer cost of the endline survey from $9.30 per farmer to $11.90 per farmer due to increased labor costs associated with hiring the plot mapping team, which operated separately from the survey team. In addition, 20 Garmin eTrex 30x devices were purchased at a cost of about USD $200 per unit for the plot mapping. Assuming each device is used in five survey rounds, on average, the cost per plot map of the devices is about $0.60. Taking into account the cost of the GPS units, the cost of adding plot mapping onto an existing survey is about $3.20 per farmer. Since an independent team of surveyors collected the plot mapping data, we estimate that the cost of plot mapping would be similar if it were conducted as a standalone exercise assuming that the surveyors had the contact information and knew the village in which each farmer's plot was located. In contrast, the field team estimates that it would have cost about $3.11 per farmer to complete a minimal endline survey in which only yield data was obtained.

In short, the marginal cost of adding plot mapping to an existing survey, or collecting plot boundary data as an independent activity, is similar to a minimal survey. Hence, using satellite data may reduce research costs by about 25% if it replaces a full-length endline survey, but not if it replaces a minimal follow-up that only obtains yield data. Satellite imagery may also allow researchers to forego additional rounds of data collection or conduct phone surveys to measure outcomes that are less noisy than yields, and therefore reduce overall survey costs. We believe that both of these benefits speak to the value of further research examining the ability of satellites to detect treatment effects in agricultural interventions.

We focus on the cost of survey data as opposed to crop cut data because the majority of agricultural development economics studies use farmer-reported data. Although crop cuts are more accurate and robust to surveyor demand ef-

[18] We are grateful to an anonymous referee that made this point.

fects, they are several times more expensive than survey data. The fact that survey data is used far more often than crop cuts suggests that economists have found surveys more effective given costs and practical constraints. Furthermore, crop cuts are subject to differential attrition, which limits their advantages compared to survey data.

The tools included in the technical appendix and online toolkit use satellite imagery and cloud computing resources that are free to use for academic research. Furthermore, the code we provide automates satellite image selection, processing, and the calculation of vegetation index values. Users are only required to upload their plot boundary data to the Google Earth Engine and select the date range and parameters that they are interested in. The program will then export a csv file containing a panel data set containing the VI measurements for each plot and each cloud free satellite image available that can be imported into any statistical software and analyzed with standard econometric methods. Hence, we expect that the estimates of field cost differences with satellite versus survey data closely reflect total cost differences since the accompanying toolkit does not require significant time investments to generate satellite yield measurements.

5 Conclusion

Increasing availability of new technologies creates opportunities to expand access to high-quality agricultural information to small-holder farmers at a low cost. Governments, practitioners, and private-sector players in the agricultural sector offer innovative solutions to generate and deliver more precise agricultural advice to farmers. Relatively little is known, however, about how such information affects farmer behavior and agricultural practices. As more resources become directed towards improving and scaling precision farming technologies for farmers in developing countries, it is critical to understand how to best deliver agricultural information to facilitate improvement in farming practices and agricultural outcomes. This study explores this question in the context of fertilizer usage in a field experiment among cotton farmers in India. We provide initial evidence that customized agricultural advice, generated based on the results of plot-level soil tests and delivered through soil health cards and mobile phones, could increase the adoption of appropriate fertilizers and improve soil fertility management practices. Farmers receiving customized fertilizer recommendations report a significantly higher likelihood of adopting recommended fertilizers, leading to a 0.082 standard deviation reduction in the fertilizer gap.

Although the intervention improved fertilizer use, we find no evidence of a change in yields. There are several reasons why the treatment may not have had an effect on agricultural productivity. First, the reported differences in fertilizer use may be due to demand effects and not true differences in fertilizer practices. We think that this explanation is unlikely because fertilizer recommendations were given in kilogram per vigha (a local area measurement), but we collected fertilizer application data in terms of kilograms and then divided reported values by plot size to construct the variables used in our analysis. Second, the null effect may be driven by unusually low rainfall, which could have inhibited nutrient absorption. It is possible that the intervention would have had a positive effect on yields in a year with normal rainfall in which crops are better able to absorb nutrients in the soil.[19] Third, increased fertilizer use may

[19] Our measures of irrigation did not differentiate between systems that depend on rainfall and those that do not depend on rainfall, preventing us from testing this hypothesis with heterogeneity analysis.

have crowded out other inputs; unfortunately we do not have sufficient data to test this hypothesis.

The lack of yield increases among treated farmers despite changes in fertilizer practices suggests the need for future research. Questions relating to the role of weather in customized fertilizer recommendations, the importance of balanced fertilizer application, and the optimal set of recommendations given soil test results lack clear answers. These topics are important given the Government of India's significant investment in the Soil Health Card Scheme. The disappointing results of this study suggest that customized soil fertility information may not always increase productivity, even when comprehension and adoption of the advice is high. Hence, further research is necessary to determine which fertilizer recommendations are effective in practice.

Finally, this study examines the potential of satellites to measure the yield effects of agricultural interventions. We find high correlations between farmer-reported and satellite-measured yields, suggesting that satellites can measure yields reliably. While linear calibration suggests that satellites can produce more precise estimates, we present evidence that OLS calibration biases treatment effect estimates. We propose the use of 2SLS calibration to solve this problem, eliminating the precision gains of satellites under random attrition in this sample. This suggests the need for more work examining the feasibility of using satellite data in RCTs, which raises econometric challenges that do not exist if one's goal is prediction and not parameter estimation.

6 Appendix I: Measuring yields with satellite data

6.1 Background

Accurately measuring agricultural yields is essential to evaluations of interventions aimed at improving outcomes for small-holder farmers. However, recent research demonstrates that farmer-reported yield data, which is used to evaluate most interventions, is very noisy and potentially biased (e.g. Carletto et al. (2015); Lobell et al. (2020)). This makes it particularly difficult to study low cost programs that have a small absolute benefit but potentially high benefit-to-cost ratio, and can lead authors to focus on intermediate outcomes such as behavior change, which in turn makes it difficult to compare the effectiveness of different intervention types.

A recent body of research has examined satellite yield measurements as an alternative to farmer-reported outcomes. Researchers have used satellite imagery to measure agricultural productivity on industrialized farms for decades, and a large body of research demonstrates that these techniques are reliable (Lobell, 2013). Studies including Jain et al. (2016), Burke and Lobell (2017), Lambert et al. (2018), Lobell et al. (2020), and Lobell et al. (2019) show that satellites can reliably measure small-holder yields, often with greater accuracy than survey data. Benami et al. (2021) note that relationships between ground and satellite measures are consistently weaker in low-income versus high-income countries, however. This suggests that satellite measures are noisier in low-income settings, so it is important to consider the role of measurement error in estimates.

Lobell et al. (2020) present strong evidence that satellites can measure smallholder crop yields. The authors find that the adjusted R^2 between satellite data and full-plot crop cut data is 0.55 on pure stand maize plots 0.1 hectares or larger in Uganda. By comparison, the adjusted R^2 between sub-plot crop cuts and full-plot crop cuts is 0.47 on these fields. Moreover, satellites are able to capture the correlation between yield and key production factors, including fertilizer use and soil quality, on both pure stand and inter-cropped plots. In contrast, farmer-reported yield data explains very little variation in full-plot crop cut data, and farmer-reported productivity is, on average, over twice as high as productivity measured using full or sub-plot crop cuts.

Jain et al. (2019) demonstrate that satellites can detect the effect of an intervention in a split-plot experiment. The authors evaluate the effect of a product to more evenly spread fertilizer on wheat plots in India using sub-plot crop cut data and satellite imagery. Results show strong agreement between the two sources of yield data at the plot ($R^2 = .55$) and sub-plot ($R^2 = .46$) level. In addition, the authors show that satellite data can reliably detect productivity gains associated with the intervention.

We expand on the satellite yield measurement literature by examining how satellite data performs in a randomized controlled trial (RCT). We make three contributions to the literature, which are detailed in Section 4.1. We repeat these contributions here for convenience.

First, we test whether the promising results of satellite measurement in the split-plot design evaluated in Jain et al. (2019) carry over to RCTs. In a split-plot design, factors — such as atmospheric interference — that may cause intra-plot noise to dominate experimental effects in an RCT are minimal because they influence each section of a plot equally. Hence, results may not directly transfer to an RCT. Furthermore, we consider how measurement error may bias treatment effect estimates calculated using satellite data calibrated via OLS regressions. We show that this is likely to lead researchers to underestimate treatment effects, and we note that the use of two-stage least squares in the calibration step produces estimates robust to measurement error, but with higher mean squared error.

Second, we explore the effects of replacing farmer-reported yield data with satellite imagery on statistical power. Satellite imagery may not offer power gains relative to farmer-reported outcomes if farmers can report yield changes accurately. Since many sources of error in farmer-reported outcomes (e.g. non-standard units and variation in harvesting time) are serially correlated, controlling for pre-intervention farmer-reported productivity may be adequate to obtain precise treatment effect estimates from survey data. We test whether satellite imagery improves statistical power relative to farmer-reported data and quantify improvements so that researchers can determine whether the data source would be cost effective in their studies. Our results show that using OLS calibration suggests that satellites can dramatically improve statistical power, but the evidence indicates that this is because attenuation bias compresses the distribution of predicted yields. If we instead use two-stage least squared to predict yields, we conclude that researchers would need a larger sample size when working with satellite data, unless differential attrition exists.

Third, we provide a "cookbook" to assist researchers without remote sensing expertise in generating satellite yield estimates and applying them to RCTs.[20] We aim to make satellite yield measurement methods evaluated in previous studies accessible to other disciplines so that the benefits of these advancements can be more widely realized. We aim to update the repository to reflect the best available methods, and we welcome other authors to contribute to it. None of the resources presented require remote sensing or coding knowledge, and they use computing resources that are free to academics when possible.

6.2 Remote sensing data

We use three sources of data to construct and analyze satellite yield measurements. First, we use GPS plot boundary data that was collected concurrently to the endline survey, but by a different survey team. Surveyors collected the plot boundary data by walking the boundaries of each farmer's plot with Garmin eTrex 30x GPS devices. Plot boundary data is available for 1,389 plots and 1,326 cotton plots. Second, we use farmer-reported plot size and cotton yield data that was collected during the endline survey. Third, we use satellite data from the Sentinel-2 constellation. We downloaded and processed low-cloud imagery from one day in 2016, three days in 2017, and five days in 2018. This paper includes a technical appendix which provides an in-depth description of each data source and information about processing of the satellite imagery.

6.3 Methodology

Satellite yield estimates were constructed using the Sentinel-2 imagery, plot boundary data, and farmer-reported productivity. We first calculated five vegetation indices (VIs) from the Sentinel-2 satellite imagery, which measure crop health and are positively related to yield. We generate separate productivity estimates from each of the VIs, and then

[20] Available at `https://github.com/gkilleen33/rs-economics`

use the best performing index to estimate treatment effects. The VIs are constructed by taking combinations of different spectral bands. The equations used to calculate each VI are presented in Appendix Table 10, and Figure 1 presents a sample true color image and Red-edge Normalized Difference Vegetation Index from a subset of the sample area.

Calculating VIs produces a single band image (for each VI) from the multispectral input images, where each pixel is a georeferenced VI measurement. Each cotton field contains multiple pixels, which we reduce to a single measurement by taking the median value of the pixels contained in the plot. This results in a panel consisting of one measurement per farmer in 2016, three measurements per farmer in 2017, and five measurements per farmer in 2018 for each VI.

We next select, for each plot, the maximum VI reading obtained in each growing season. This approach reduces the influence of sowing time, though in practice it makes virtually no difference, because cloud cover was not an issue, and almost the entire sample was imaged at the same time.[21] In other settings, we expect that taking the maximum VI value across multiple satellite passes could substantially improve the accuracy of results and statistical power.

We assess whether satellite imagery can reliably measure small-holder cotton yields in this sample by regressing the maximum VI values on farmer-reported productivity. We estimate the OLS regression

$$VI_i = \alpha + \beta Yield_i + \epsilon_i$$

where VI_i is the maximum vegetation index value observed in plot i in 2018 and $Yield_i$ is 2018 farmer-reported yield in metric tons per hectare. We choose a linear form because Stamatiadis et al. (2010) find a linear relationship between NDVI and cotton yields. Details about the software packages required to construct satellite yield estimates and commented, adaptable code are available in the technical appendix and at `https://github.com/gkilleen33/rs-economics`. These resources do not require remote sensing or programming knowledge to use.

We construct satellite yield estimates to use in treatment effect calculations from the best performing VI, which is shown in Section 4.1 to be the Red-edge Normalized Difference Vegetation Index (reNDVI, see Appendix Table 9 for more information). To convert the VI values to yield values, we estimate the regression

$$Yield_i = \alpha + \beta reNDVI_i + \epsilon_i$$

where $Yield_i$ is farmer-reported total harvest in kilograms over GPS measured plot size in hectares and $reNDVI_i$ is the maximum reNDVI value observed in plot i. As discussed in the main paper, we estimate this regression – referred to as the calibration step – using both OLS and two-stage least squares in which we instrument for $reNDVI_i$ using sowing date and total rainfall from June-October. We then linearly fit the VI values to yield using the coefficients estimated from the regression.

We compare the utility of satellite yield measurements to farmer-reported data by estimating treatment effects using both data sources. Power gains from satellite imagery may come from at least three sources: more precise

[21] Sentinel-2 satellites collect data in swaths that are roughly North-South and 290 km wide. If the sample crosses a swath boundary, then Sentinel-2 will collect images of different parts of the sample on different days.

measurements, reduced attrition, and multiple years of pre-intervention productivity data. If differential attrition is a concern, then satellite imagery may also improve power because differential attrition can be eliminated. We estimate treatment effects using satellite data with two specifications to disentangle the power benefits of more precise measurements and multiple years of pre-intervention productivity data. The power benefits of reduced attrition and differential attrition are examined in power calculations detailed in Section 4.4.

First, we estimate

$$Y_{2018,i}^{Sat} = \alpha_b + \beta T_i + \theta_1 reNDVI_{2017,i}^{Sat} + \epsilon_i$$

where $Y_{2018,i}^{Sat}$ denotes post-intervention satellite measured yield for individual i, α_b denotes block fixed effects, T_i denotes treatment, and $reNDVI_{2017,i}^{Sat}$ is the 2017 Red-edge NDVI measurement, which is equivalent to controlling for 2017 yield predictions since yield is an affine transformation of reNDVI.

Second, we consider

$$Y_{2018,i}^{Sat} = \alpha_b + \beta T_i + \theta_1 reNDVI_{2017,i}^{Sat} + \theta_2 reNDVI_{2016,i} + \epsilon_i$$

where $reNDVI_{2016,i}$ is the reNDVI value observed in individual i's plot in the single 2016 Sentinel-2 image that we downloaded.

References

Lori Beaman, Dean Karlan, Bram Thuysbaert, and Christopher Udry. Profitability of fertilizer: experimental evidence from female rice farmers in Mali. *American economic review*, 103(3):381–386, 5 2013. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.3.381.

Elinor Benami, Zhenong Jin, Michael R Carter, Aniruddha Ghosh, Robert J Hijmans, Andrew Hobbs, Benson Kenduiywo, and David B Lobell. Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews Earth & Environment*, 2(2):140–159, 2021. ISSN 2662-138X. doi: 10.1038/s43017-020-00122-y. URL https://doi.org/10.1038/s43017-020-00122-y.

P. Bonhard and M. A. Sasse. 'Knowing me, knowing you' - Using profiles and social networking to improve recommender systems. *BT Technology Journal*, 24(3):84–98, 7 2006. ISSN 13583948. doi: 10.1007/s10550-006-0080-3.

Marshall Burke and David B Lobell. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, 114(9):2189–2194, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1616919114. URL https://www.pnas.org/content/114/9/2189.

Calogero Carletto, Dean Jolliffe, and Raka Banerjee. From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *The Journal of Development Studies*, 51(2):133–148, 2015. doi: 10.1080/00220388.2014.968140. URL https://doi.org/10.1080/00220388.2014.968140.

Lorenzo Casaburi, Michael Kremer, Sendhil Mullainathan, and Ravindra Ramrattan. Harnessing ICT to Increase Agricultural Production: Evidence from Kenya. 2019. URL https://scholar.harvard.edu/kremer/publications/harnessing-ict-increase-agricultural-productionevidence-kenya.

Kenneth G Cassman, Achim Dobermann, and Daniel T Walters. Agroecosystems, Nitrogen-Use Efficiency, and Nitrogen Management. *Ambio*, 31(2):132–140, 2002. ISSN 00447447, 16547209. URL www.jstor.org/stable/4315226.

Shawn Cole and A. Nilesh Fernando. 'Mobile'izing Agricultural Advice: Technology Adoption, Diffusion, and Sustainability. 2020. URL https://www.dropbox.com/s/m68t6h46q7e169z/ao_032020.pdf?dl=0.

Shawn Cole and Garima Sharma. The Promise and Challenges of Implementing ICT in Indian Agriculture. 2017.

J Dash and P J Curran. The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25(23):5403–5413, 2004. doi: 10.1080/0143116042000274015. URL https://doi.org/10.1080/0143116042000274015.

Esther Duflo, Michael Kremer, and Jonathan Robinson. How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. *American Economic Review*, 98(2):482–488, 5 2008. doi: 10.1257/aer.98.2.482. URL https://www.aeaweb.org/articles?id=10.1257/aer.98.2.482.

Raissa Fabregas, Michael Kremer, Matthew Lowes, Robert On, and Giulia Zane. SMS-extension and Farmer Behavior: Lessons from Six RCTs in East Africa. 2019.

Ram Fishman, Avinash Kishore, Yoav Rothler, and Patrick Ward. Can Information Help Reduce Imbalanced Application of Fertilizers in India? Experimental Evidence from Bihar. 2016. doi: 10.22004/AG.ECON.235705.

Anatoly A Gitelson, Andrés Viña, Timothy J Arkebauer, Donald C Rundquist, Galina Keydan, and Bryan Leavitt. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, 30 (5), 2003. doi: 10.1029/2002GL016450. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL016450.

R Haas, J Schell, and D Deering. Monitoring vegetation systems in the Great Plains with ERTS. In *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A p 309-317 (SEE N74-30705 20-13); United States*, volume 1, 1974. URL http://search.proquest.com/docview/22161195/.

Aurélie P. Harou, Malgosia Madajewicz, Hope Michelson, Cheryl A. Palm, Nyambilila Amuri, Christopher Magomba, Johnson M. Semoka, Kevin Tschirhart, and Ray Weil. The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania. *Journal of Development Economics*, 155: 102707, 3 2022. ISSN 0304-3878. doi: 10.1016/J.JDEVECO.2021.102707.

Meha Jain, Amit Srivastava, Balwinder-Singh, Rajiv Joon, Andrew McDonald, Keitasha Royal, Madeline Lisaius, and David Lobell. Mapping Smallholder Wheat Yields and Sowing Dates Using Micro-Satellite Data. *Remote Sensing*, 8(10):860, 10 2016. ISSN 2072-4292. doi: 10.3390/rs8100860. URL `http://www.mdpi.com/2072-4292/8/10/860`.

Meha Jain, Balwinder-Singh, Preeti Rao, Amit K Srivastava, Shishpal Poonia, Jennifer Blesh, George Azzari, Andrew J McDonald, and David B Lobell. The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustainability*, 2(10):931–934, 2019. ISSN 2398-9629. doi: 10.1038/s41893-019-0396-x. URL `https://doi.org/10.1038/s41893-019-0396-x`.

W James and Charles Stein. Estimation with Quadratic Loss. In Samuel Kotz and Norman L Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 443–460. Springer New York, New York, NY, 1992. ISBN 978-1-4612-0919-5. doi: 10.1007/978-1-4612-0919-5{_}30. URL `https://doi.org/10.1007/978-1-4612-0919-5_30`.

Harmandeep S Khurana, Steven B Phillips, Bijay-Singh, Achim Dobermann, Ajmer S Sidhu, Yadvinder-Singh, and Shaobing Peng. Performance of Site-Specific Nutrient Management for Irrigated, Transplanted Rice in Northwest India. *Agronomy Journal*, 99(6):1436–1447, 2007. URL `https://acsess.onlinelibrary.wiley.com/doi/abs/10.2134/agronj2006.0283`.

Marie-Julie Lambert, Pierre C Sibiry Traoré, Xavier Blaes, Philippe Baret, and Pierre Defourny. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. *Remote Sensing of Environment*, 216:647 – 657, 2018. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2018.06.036. URL `http://www.sciencedirect.com/science/article/pii/S0034425718303146`.

David S Lee. Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies*, 76(3):1071–1102, 12 2009. ISSN 0034-6527. doi: 10.1111/j.1467-937X.2009.00536.x. URL `https://doi.org/10.1111/j.1467-937X.2009.00536.x`.

Jeffrey B Lewis and Drew A Linzer. Estimating Regression Models in Which the Dependent Variable Is Based on Estimates. *Political Analysis*, 13(4):345–364, 2005. ISSN 10471987, 14764989. URL `http://www.jstor.org/stable/25791822`.

David B Lobell. The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143:56–64, 2013. ISSN 03784290. doi: 10.1016/j.fcr.2012.08.008. URL `http://www.sciencedirect.com/science/article/pii/S0378429012002754`.

David B. Lobell, Stefania Di Tommaso, Calum You, Ismael Yacoubou Djima, Marshall Burke, and Talip Kilic. Sight for Sorghums: Comparisons of Satellite- and Ground-Based Sorghum Yield Estimates in Mali. *Remote Sensing*, 12 (1):100, 12 2019. ISSN 2072-4292. doi: 10.3390/rs12010100. URL `https://www.mdpi.com/2072-4292/12/1/100`.

David B Lobell, George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray. Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *American Journal of Agricultural Economics*, 102(1):202–219, 1 2020. ISSN 0002-9092. doi: https://doi.org/10.1093/ajae/aaz051. URL `https://doi.org/10.1093/ajae/aaz051`.

Wang Maohua. Possible adoption of precision agriculture for developing countries at the threshold of the new millennium. *Computers and Electronics in Agriculture*, 30(1-3):45–50, 2 2001. ISSN 01681699. doi: 10.1016/S0168-1699(00)00154-X.

P A Matson, R Naylor, and I Ortiz-Monasterio. Integration of environmental, agronomic, and economic aspects of fertilizer management. *Science (Washington)*, 280(5360):112–115, 4 1998. URL `https://science.sciencemag.org/content/280/5360/112`.

David McKenzie. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, 99(2):210 – 221, 2012. ISSN 0304-3878. doi: https://doi.org/10.1016/j.jdeveco.2012.01.002. URL http://www.sciencedirect.com/science/article/pii/S030438781200003X.

Pinaki Mondal and Manisha Basu. Adoption of precision agriculture technologies in India and in some developing countries: Scope, present status and strategies, 6 2009. ISSN 10020071.

U Mueller-Wilm. Sen2Cor Configuration and User Manual, 2019. URL http://step.esa.int/thirdparties/sen2cor/2.8.0/docs/S2-PDGS-MPC-L2A-SUM-V2.8.pdf.

Agustin Pagani, John Sawyer, and Antonio Mallarino. Site-Specific Nutrient Management: For Nutrient Management Planning To Improve Crop Production, Environmental Quality, and Economic Return. *Extension and Outreach Publications*, 5 2013. URL https://lib.dr.iastate.edu/extension_pubs/116.

M. F. Pampolino, I. J. Manguiat, S. Ramanathan, H. C. Gines, P. S. Tan, T. T.N. Chi, R. Rajendran, and R. J. Buresh. Environmental impact and economic benefits of site-specific nutrient management (SSNM) in irrigated rice systems. *Agricultural Systems*, 93(1-3):1–24, 3 2007. ISSN 0308521X. doi: 10.1016/j.agsy.2006.04.002.

Jonathan Proctor, Tamma Carleton, and Sandy Sum. Parameter Recovery Using Remotely Sensed Variables. *National Bureau of Economic Research Working Paper Series*, No. 30861, 2023. doi: 10.3386/w30861. URL http://www.nber.org/papers/w30861http://www.nber.org/papers/w30861.pdf.

David Schimmelpfennig. Farm Profits and Adoption of Precision Agriculture. (1477-2016-121190):46, 2016. doi: 10.22004/ag.econ.249773. URL http://ageconsearch.umn.edu/record/249773.

Elizabeth Sillence, Pam Briggs, Peter Harris, and Lesley Fishwick. A framework for understanding trust factors in web-based health advice. *International Journal of Human - Computer Studies*, 64(8):697–713, 2006. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2006.02.007.

Stamatis Stamatiadis, Christos Tsadilas, and James S Schepers. Ground-based canopy sensing for detecting effects of water stress in cotton. *Plant and Soil*, 331(1):277–287, 6 2010. doi: 10.1007/s11104-009-0252-2. URL https://doi.org/10.1007/s11104-009-0252-2.

Tavneet Suri. SELECTION AND COMPARATIVE ADVANTAGE IN TECHNOLOGY ADOPTION. *Econometrica*, 79(1):159–209, 2011. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/41057440.

Hussain R. Usman, Mohammad H. Rahbar, Sibylle Kristensen, Sten H. Vermund, Russell S. Kirby, Faiza Habib, and Eric Chamot. Randomized controlled trial to improve childhood immunization adherence in rural Pakistan: redesigned immunization card and maternal education. *Tropical Medicine & International Health*, 16(3):334–342, 3 2011. ISSN 13602276. doi: 10.1111/j.1365-3156.2010.02698.x. URL http://doi.wiley.com/10.1111/j.1365-3156.2010.02698.x.

Andrés Viña and Anatoly A Gitelson. New developments in the remote estimation of the fraction of absorbed photosynthetically active radiation in crops. *Geophysical Research Letters*, 32(17), 2005. doi: 10.1029/2005GL023647. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL023647.

M Weiss and F Baret. Evaluation of Canopy Biophysical Variable Retrieval Performances from the Accumulation of Large Swath Satellite Data. *Remote Sensing of Environment*, 70(3):293 – 306, 1999. ISSN 0034-4257. doi: https://doi.org/10.1016/S0034-4257(99)00045-0. URL http://www.sciencedirect.com/science/article/pii/S0034425799000450.

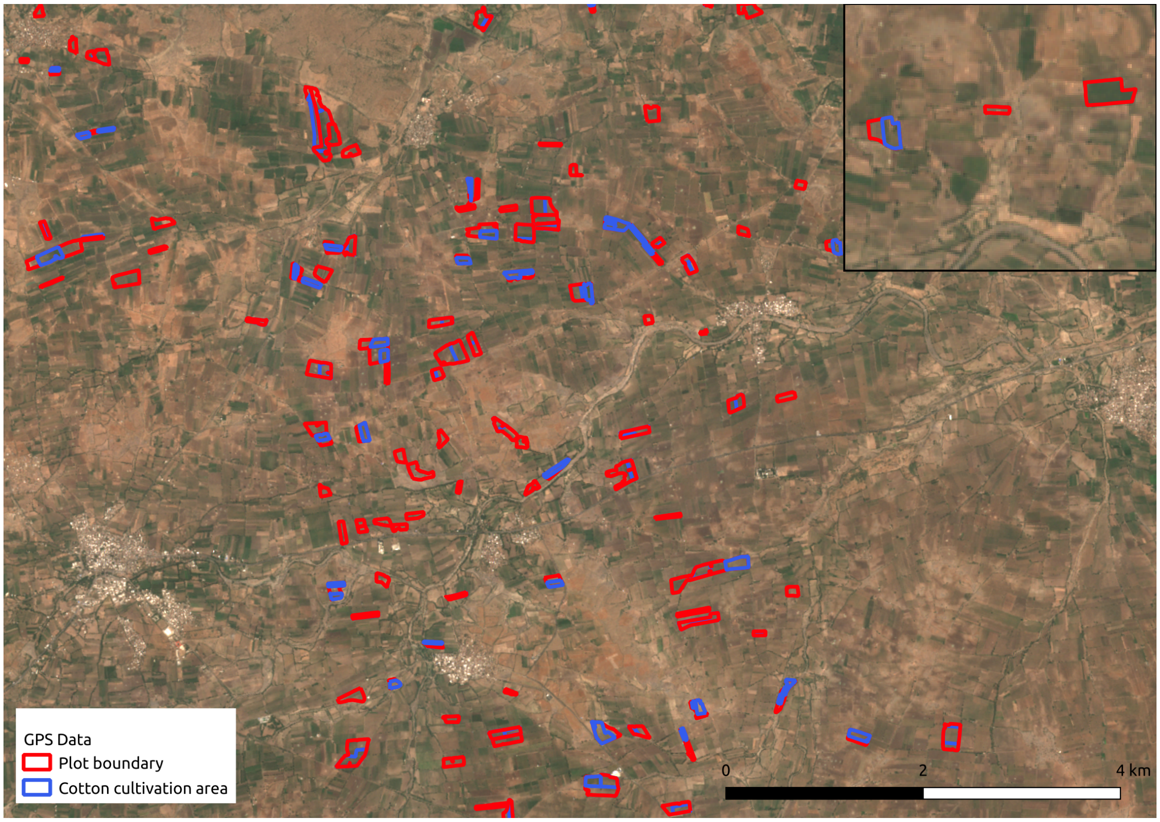
Marie Weiss and Fred Baret. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER, 2016. URL https://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf.

Duli Zhao, K Raja Reddy, V Gopal Kakani, John J Read, and Sailaja Koti. Canopy reflectance in cotton for growth assessment and lint yield prediction. *European Journal of Agronomy*, 26(3):335 – 344, 2007. ISSN 1161-0301. doi: https://doi.org/10.1016/j.eja.2006.12.001. URL http://www.sciencedirect.com/science/article/pii/S1161030106001626.

7 Figures

Figure 1: Example Sentinel-2 imagery: October 28th, 2018

(a) True color composite



(b) reNDVI

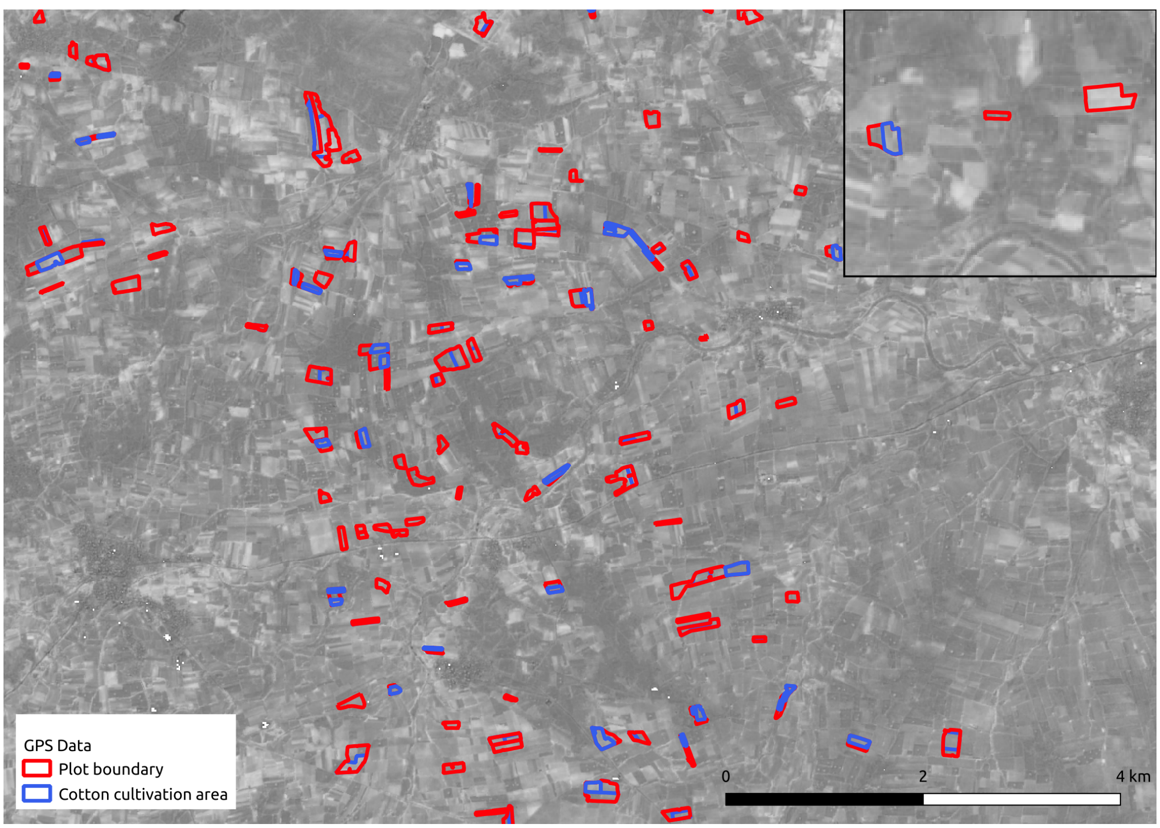


Figure 1 presents sample Sentinel-2 imagery collected on October 28th, 2018 in a portion of the sample region that was selected due to a high density of sample plots. Panel (a) presents a true color composite. Panel (b) presents a reNDVI image, which is the vegetation index used for satellite productivity measurements in this paper. Polygons in red represent the boundaries of sample plots. Polygons in blue show the area on which the farmer cultivated cotton, if they did not cultivate cotton on their full plot.

Figure 2: Applied minus recommended fertilizer, basal dose (kg/ha)

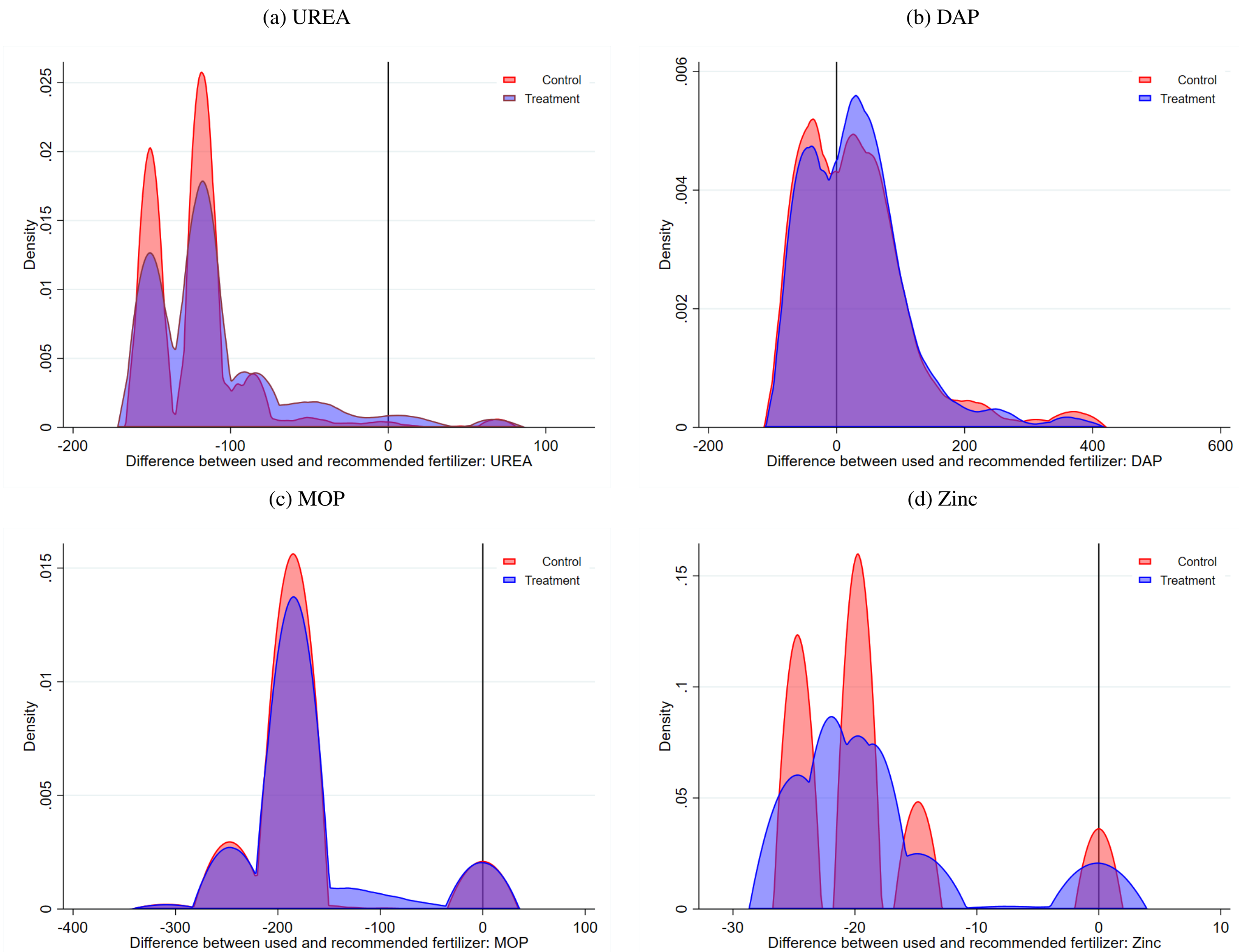


Figure 2 displays kernel density estimates for the difference between the applied and recommended basal fertilizer dose in kg/ha. Differences were calculated by subtracting the lab recommended fertilizer dose from the farmer's self-reported use. The density plots use a Epanechnikov kernel function. Estimates were calculated separately for treatment (red) and control (blue). Values below 0 indicate that the farmer applied less than the recommended dose, and values above 0 indicate that the farmer applied more than the recommended dose. Differences are winsorized at the 99th percentile.

Figure 3: Applied minus recommended fertilizer, total (kg/ha)

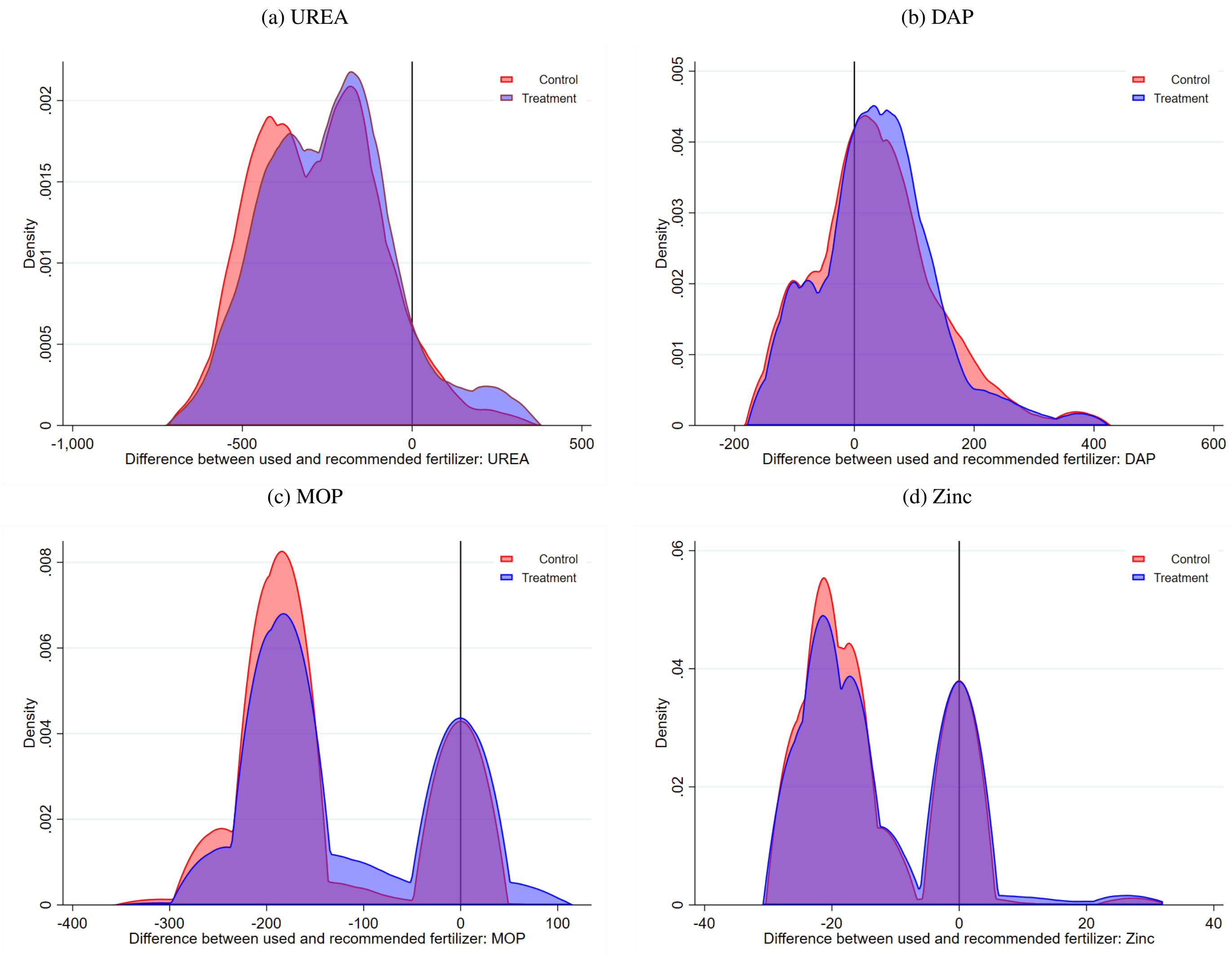


Figure 3 displays kernel density estimates for the difference between the applied and recommended total fertilizer application (across the full growing season) in kg/ha. Differences were calculated by subtracting the lab recommended fertilizer dose from the farmer's self-reported use. The density plots use a Epanechnikov kernel function. Estimates were calculated separately for treatment (red) and control (blue). Values below 0 indicate that the farmer applied less than the recommended amount, and values above 0 indicate that the farmer applied more than the recommended amount. Differences are winsorized at the 99th percentile.

Figure 4: Cotton yields (kg/ha)

(a) Survey data

(b) Satellite data

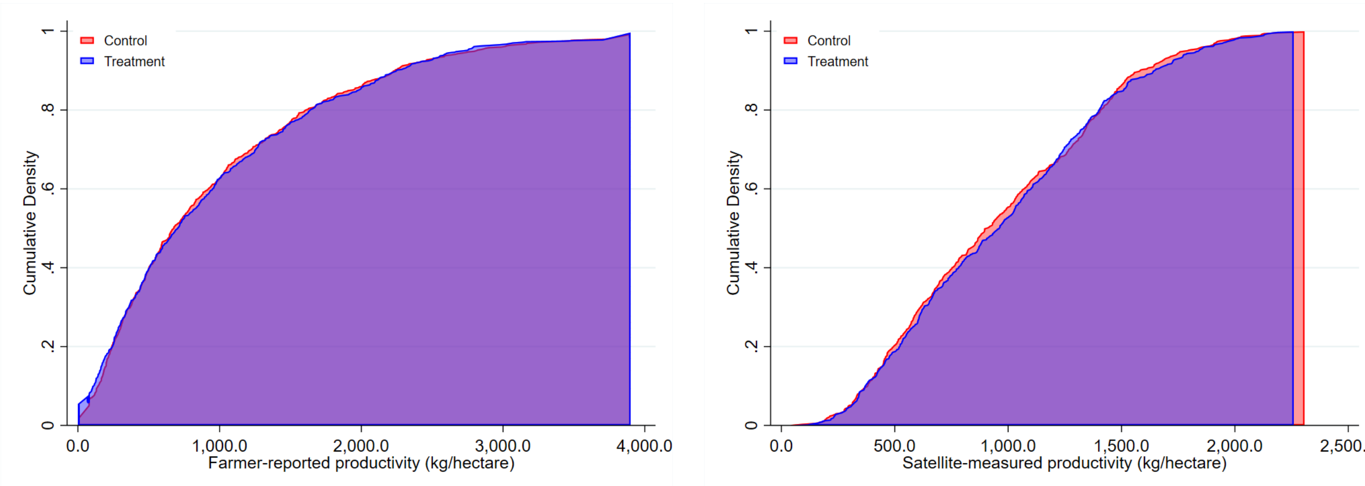
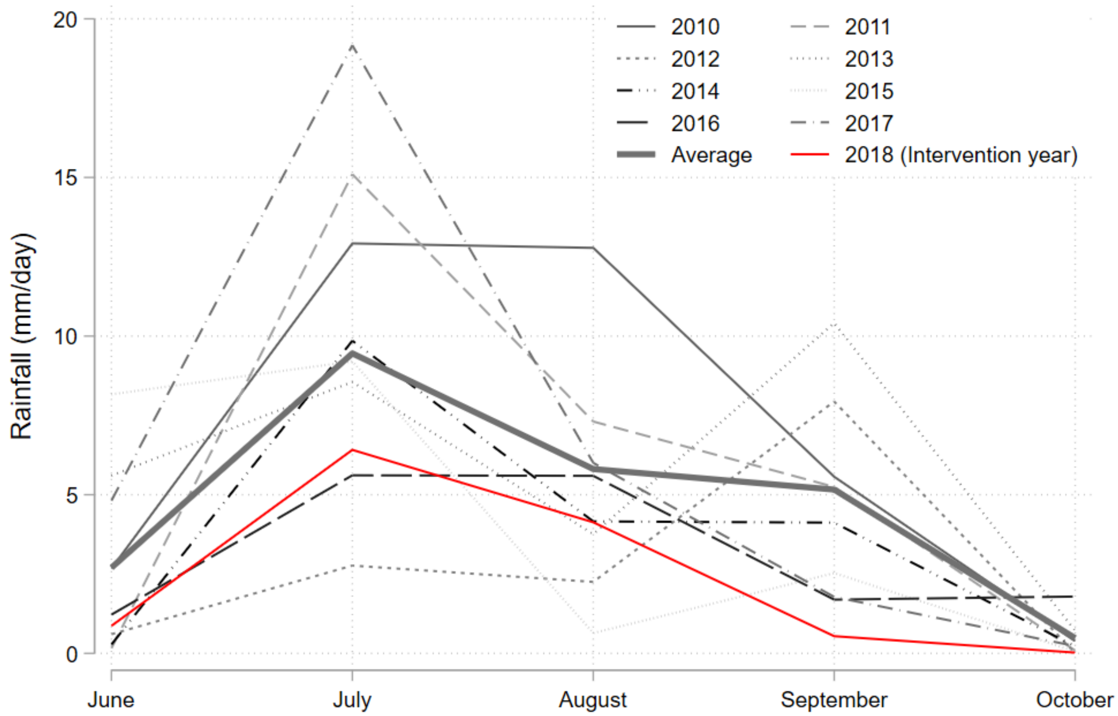


Figure 4 displays cumulative density plots of productivity in kg/ha. Panel (a) uses farmer-reported total yields divided by the GPS-measured area on which the farmer grew cotton. Positive yield values are winsorized at the 2nd and 98th percentiles. Panel (b) displays satellite measured yields based on a reNDVI (Viña and Gitelson, 2005) calculated from Sentinel-2 L2A imagery. We constructed satellite yield measurements by calculating the median value of reNDVI values contained within each plot in the sample on 5 dates and then taking the maximum reNDVI across the 5 images for each plot. This value was then linearly fitted to farmer-reported yield data. Several linear predictions were negative and were recoded to 0.

Figure 5: Rainfall

(a) Monthly



(b) Daily

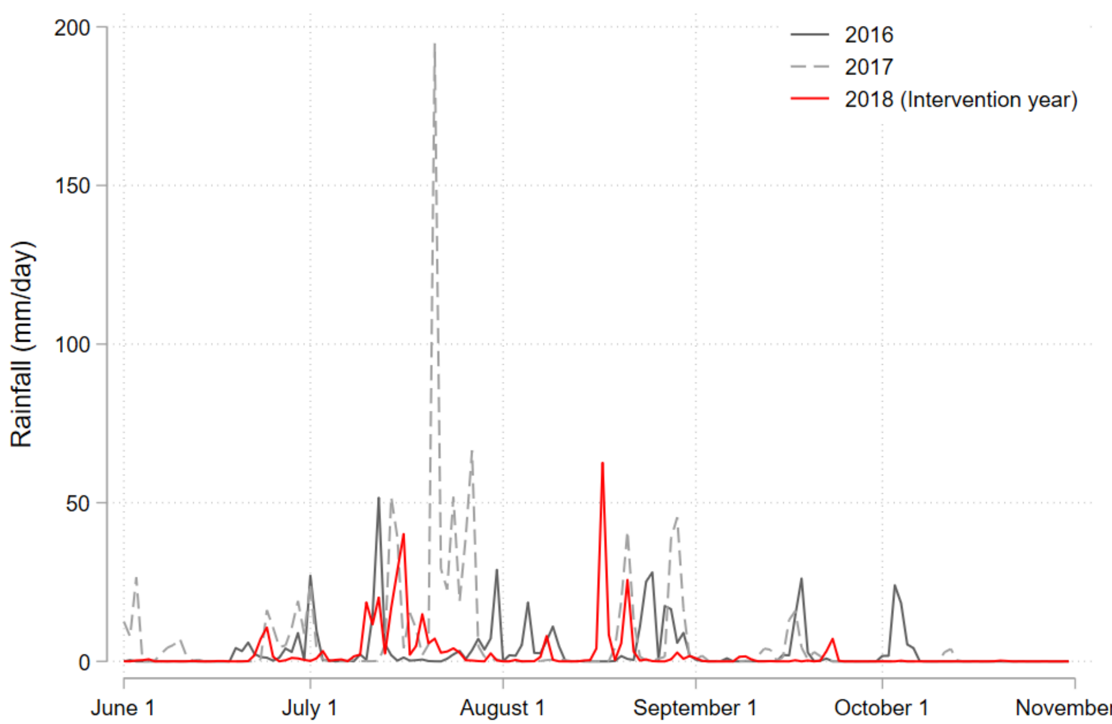
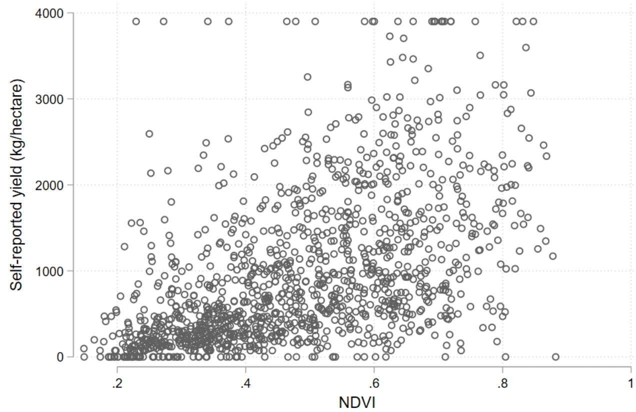


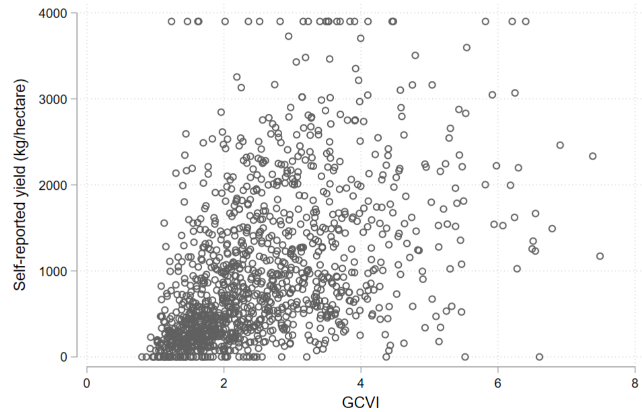
Figure 5 plots average rainfall across the sample. Rainfall data was obtained from NASA's Global Participation Measurement mission. The figures use final run IMERG data accessed through the Google Earth Engine. Sub-plot (a) presents monthly rainfall values, in mm/day, from 2010-2019. The average value from 2010-2019 is also reported. Sub-plot (b) displays daily rainfall data from 2016-2019.

Figure 6: Sentinel-2 vegetation indices vs farmer-reported yield (kg/ha)

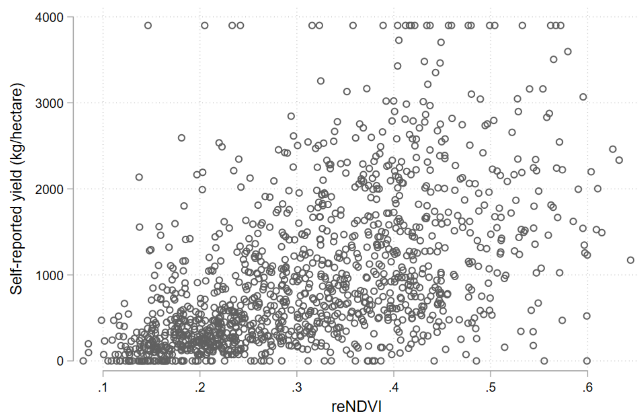
(a) NDVI



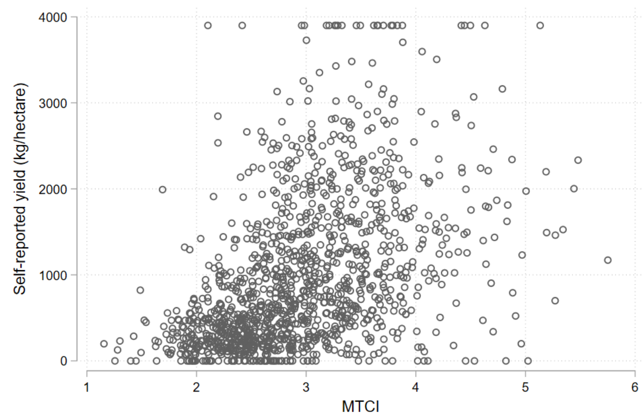
(b) GCVI



(c) reNDVI



(d) MTCI



(e) LAI

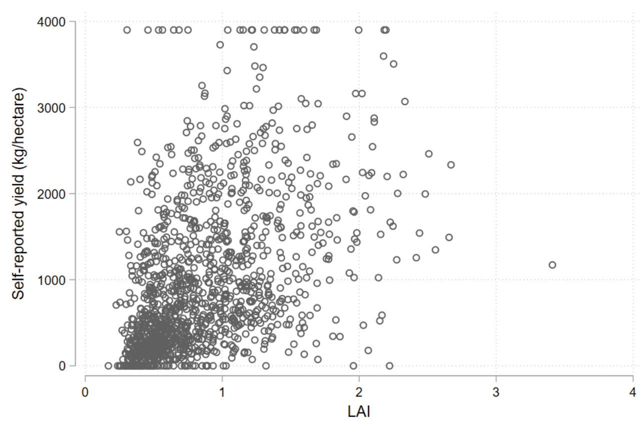
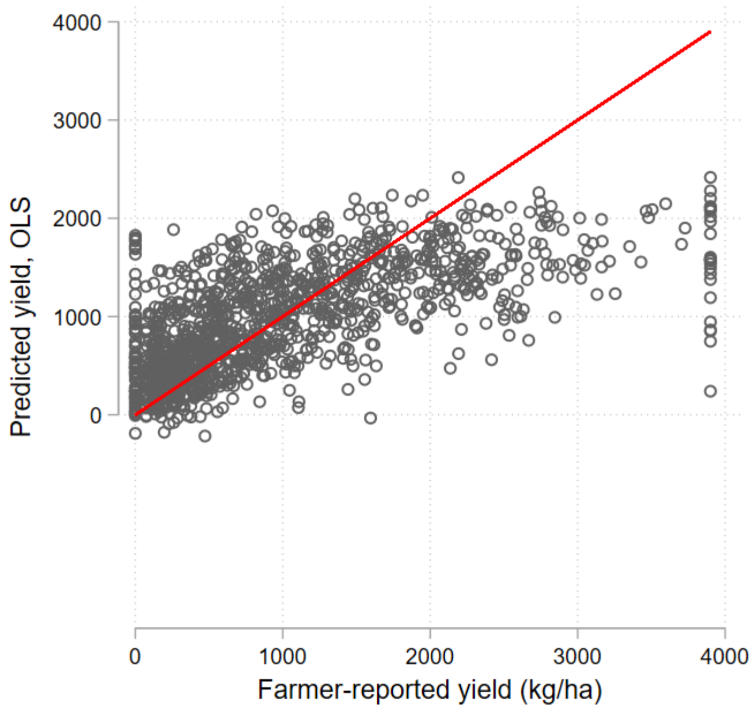


Figure 6 plots vegetation index (VI) values against farmer-reported productivity. We calculated VI values by taking the median value of each VI pixel contained in each sample plot for 5 Sentinel-2 images from 2018. We then took the maximum value across the 5 satellite images. Farmer-reported yield was calculated by taking farmer-reported total harvest data and dividing it by GPS-measured plot area. Positive farmer-reported yield values were winsorized at the 2nd and 98th percentiles.

Figure 7: Farmer-reported yields vs satellite predictions:
OLS and 2SLS calibration

(a) OLS calibration



(b) 2SLS calibration

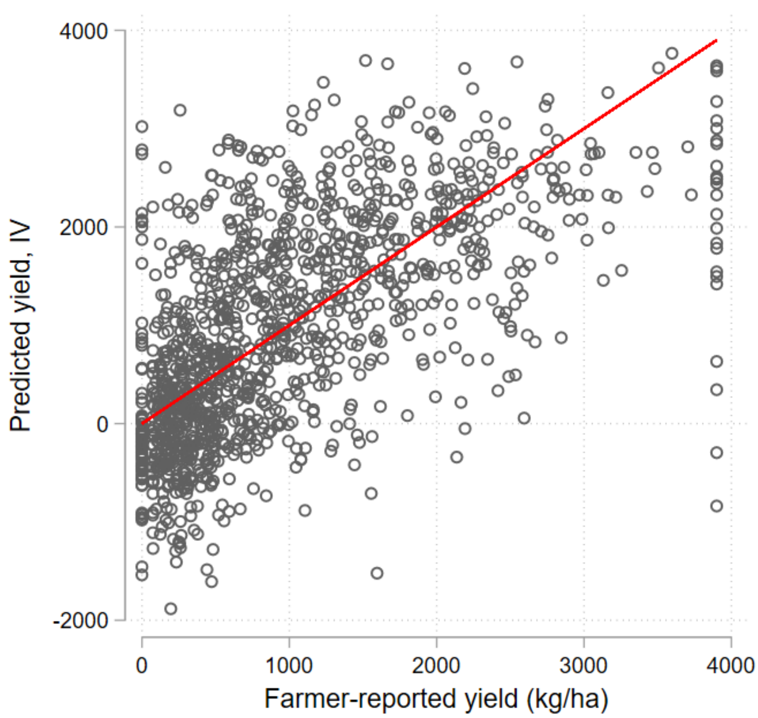


Figure 7 plots farmer-reported yields against satellite yield predictions calibrated using an OLS first stage and a 2SLS first stage. The red line is at 45 degrees. Farmer-reported yield was calculated by taking farmer-reported total harvest data and dividing it by GPS-measured plot area. Positive farmer-reported yield values were winsorized at the 2nd and 98th percentiles.

Figure 8: Treatment effect confidence intervals
Farmer-reported vs satellite measurements

(a) Monthly

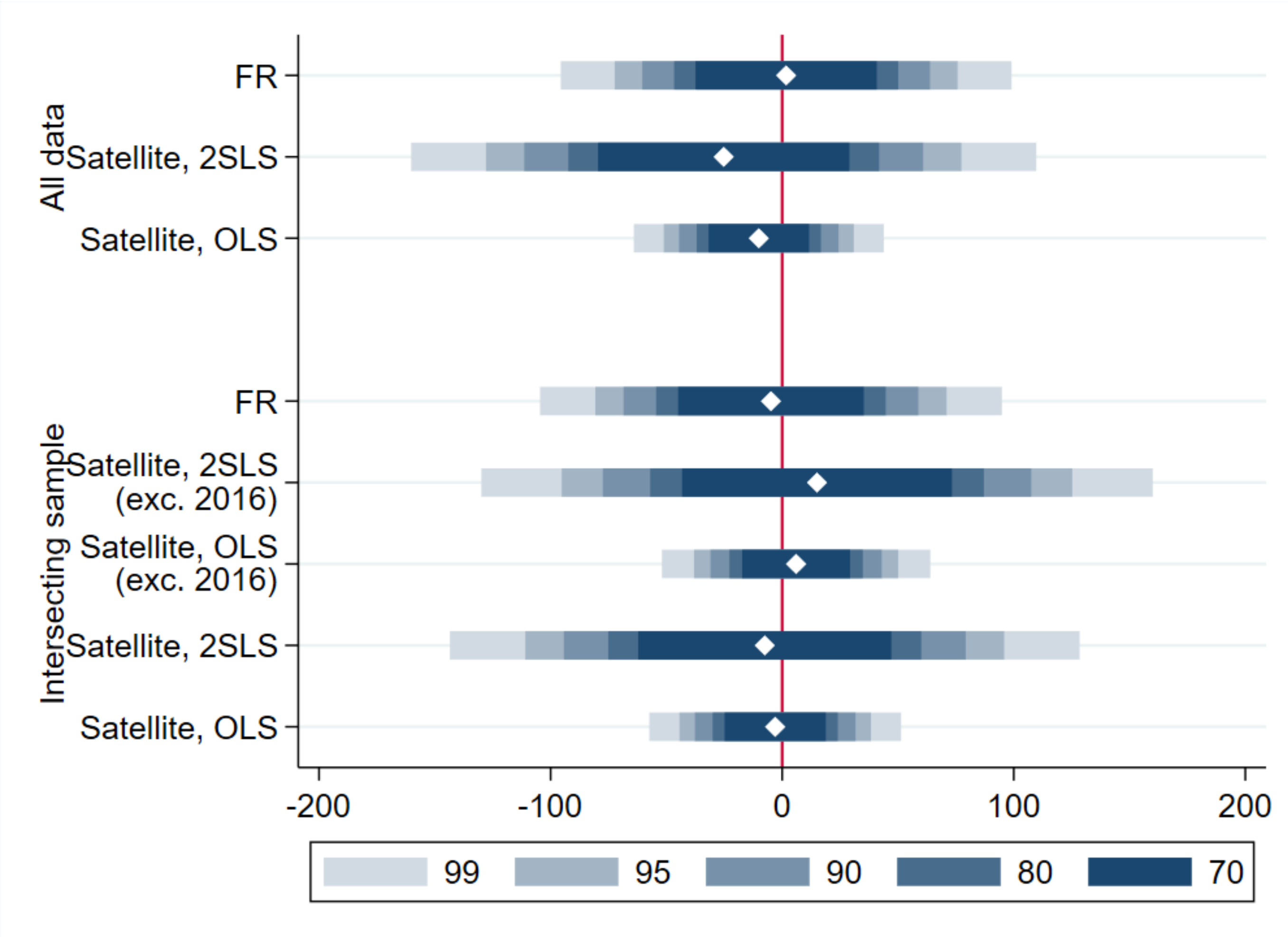


Figure 8 plots confidence intervals of the treatment effect on productivity using farmer-reported and satellite yield data. The first three rows use all available data, and the bottom five rows examine a restricted sample of data that is not missing for either data source. Rows 1 and 4 examine farmer-reported yield data, and each of the other rows use satellite measurements. Rows 1, 4, 6 and 7 control for 2017 productivity. The remaining rows control for 2016 and 2017 productivity. Rows labelled "satellite" use satellite predictions with 2SLS calibration, and rows labeled "Satellite, OLS" examine satellite yield predictions with OLS calibration. All regressions include block fixed effects and robust standard errors.

Tables

Table 1: Summary statistics, baseline

	(1) Control (full sample)	(2) Treatment - Control (full sample)	(3) T-C (basal)	(4) T-C (midline)	(5) T-C (endline)	(6) T-C (farmer-reported yield)	(7) T-C (satellite data)
Age (baseline)	42.600	-0.155	-0.464	0.036	-0.073	0.020	-0.153
	[11.485]	(0.604)	(0.633)	(0.611)	(0.623)	(0.637)	(0.643)
Literate	0.845	-0.017	-0.016	-0.023	-0.024	-0.028	-0.024
	[0.362]	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)
Total cotton land (2017)	3.602	-0.189	-0.124	-0.180	-0.197	-0.236	-0.090
	[3.735]	(0.187)	(0.201)	(0.189)	(0.191)	(0.196)	(0.184)
Sampled plot size (2017)	2.080	-0.075	-0.040	-0.089	-0.103	-0.111	-0.073
	[1.677]	(0.082)	(0.087)	(0.081)	(0.083)	(0.085)	(0.085)
Irrigation	0.911	-0.024	-0.016	-0.030*	-0.022	-0.031*	-0.016
	[0.285]	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Strong house	0.613	0.018	0.025	0.020	0.010	0.004	0.003
	[0.487]	(0.025)	(0.026)	(0.025)	(0.026)	(0.026)	(0.027)
Own plough	0.420	0.009	-0.008	0.012	0.007	0.001	0.016
	[0.494]	(0.025)	(0.027)	(0.026)	(0.026)	(0.027)	(0.027)
Crop insurance	0.570	-0.008	-0.007	-0.005	-0.009	-0.009	-0.021
	[0.495]	(0.025)	(0.027)	(0.026)	(0.026)	(0.027)	(0.027)
Children	2.415	-0.101	-0.133*	-0.090	-0.104	-0.093	-0.107
	[1.346]	(0.068)	(0.070)	(0.069)	(0.071)	(0.073)	(0.073)
$>$ median education	0.377	-0.015	-0.006	-0.019	-0.018	-0.017	-0.019
	[0.485]	(0.025)	(0.026)	(0.025)	(0.026)	(0.026)	(0.026)
Soil tested prior to study	0.142	-0.016	-0.016	-0.013	-0.017	-0.021	-0.012
	[0.349]	(0.018)	(0.019)	(0.018)	(0.018)	(0.019)	(0.019)
UREA last season (kg/ha)	292.328	-19.055	-5.956	-20.435	-23.589	-27.914	-23.270
	[444.784]	(17.328)	(9.115)	(17.794)	(18.552)	(19.225)	(19.568)
DAP last season (kg/ha)	156.450	-7.555	-8.901	-8.490	-8.777	-10.367	-8.882
	[207.443]	(9.747)	(10.601)	(10.009)	(10.418)	(10.852)	(10.977)
MOP last season (kg/ha)	5.293	0.332	0.292	0.020	-0.024	-0.100	0.109
	[23.010]	(1.215)	(1.303)	(1.226)	(1.278)	(1.309)	(1.342)
Zinc last season (kg/ha)	1.264	0.985	1.001	0.956	0.505	0.444	0.648
	[6.557]	(0.632)	(0.696)	(0.647)	(0.400)	(0.413)	(0.418)
Observations	755	1,516	1,375	1,469	1,402	1,341	1,323
p-value of joint orthogonality		0.530	0.696	0.418	0.566	0.436	0.643

Standard deviations in brackets. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column (1) reports the control group mean of the indicated variable across all baseline respondents. Column (2) reports the difference in means between the control and treatment groups across the full sample of baseline respondents. Columns (3), (4), and (5) report the same value, but among respondents that completed the basal survey, midline survey, and endline survey. Column (6) restricts the sample to observations for which farmer-reported yields are non-missing. Column (7) restricts the sample to observations for which satellite yield measurements are available for 2016, 2017, and 2018. Education data was missing for 175 observations. Number of children is missing for 25 observations. Crop insurance is missing for 20 observations. UREA and DAP usage last season are missing for 3 observations, DAP and MOP usage last season are missing for two observations, and sampled plot size is missing for one observations. We interpolated the missing values of these variables using the median value of each variable.

Table 2: Treatment effect on KT Call Engagement, and Knowledge

	Administrative data			Endline survey		Midline survey
	(1) KT call pickup rate	(2) Share of total content heard	(3) KT call rating	(4) Use mobile phone advisory	(5) Trust in mobile phone advisory (1-5)	(6) Fertilizer questions correct
Panel A: Full sample						
Treatment	-0.038***	-0.006	0.148**	0.033**	0.206***	0.266***
	(0.011)	(0.014)	(0.069)	(0.014)	(0.065)	(0.103)
Observations	1,516	1,516	1,368	1,402	1,281	1,484
Adjusted R^2	0.012	0.006	0.027	0.014	0.045	0.050
Control mean	0.864	0.554	3.757	0.906	3.267	3.210
Panel B: Respondents that provided yield data						
Treatment	-0.035***	-0.005	0.125*	0.026*	0.206***	0.309***
	(0.011)	(0.014)	(0.072)	(0.014)	(0.066)	(0.107)
R^2	0.018	0.012	0.026	0.015	0.050	0.053
Observations	1,341	1,341	1,247	1,341	1,228	1,339
Adjusted R^2	0.014	0.008	0.022	0.011	0.045	0.049
Control mean	0.870	0.565	3.785	0.911	3.282	3.222
Panel C: Respondents for which satellite data is available						
Treatment	-0.033***	-0.003	0.151**	0.036**	0.245***	0.316***
	(0.012)	(0.014)	(0.072)	(0.015)	(0.066)	(0.108)
R^2	0.017	0.014	0.030	0.020	0.052	0.054
Observations	1,326	1,326	1,216	1,325	1,211	1,314
Adjusted R^2	0.013	0.010	0.025	0.016	0.047	0.049
Control mean	0.865	0.565	3.798	0.906	3.262	3.188
Panel D: Respondents for which farmer-reported and satellite yields are available						
Treatment	-0.035***	-0.005	0.125*	0.026*	0.206***	0.309***
	(0.011)	(0.014)	(0.072)	(0.014)	(0.066)	(0.107)
R^2	0.018	0.012	0.026	0.015	0.050	0.053
Observations	1,341	1,341	1,247	1,341	1,228	1,339
Adjusted R^2	0.014	0.008	0.022	0.011	0.045	0.049
Control mean	0.870	0.565	3.785	0.911	3.282	3.222

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable in column (1) is the share of non-fertilizer KT calls that the farmer picked up. In column (2), the dependent variable is the average share of non-fertilizer KT calls that the farmer listened to. Column (3) is the ranking given by farmers to KT calls in terms of usefulness, where 1 means that the calls were not useful and 5 that the call were most useful. Column (4) examines whether farmers use a mobile phone advisory service to make agricultural decisions. In Column (5), the dependent variable records the respondent's reported trust in mobile phone-based advice on a scale of 1 (very low trust) to 5 (very high trust). Column (6) records the number of questions, out of 14, in a quiz aimed to assess fertilizer knowledge were answered correctly. The sample size is lower in column (2) than in columns (1) and (3) because not all users leave ratings. The sample size differs in columns (4), (5) and (6) because they use survey data, whereas columns (1) - (4) use administrative data. Panel (a) includes the full sample. Panel (b) is restricted to respondents that provided 2018 cotton yield data and plot size information. Panel (c) is restricted to respondents for which we have 2018 satellite yield data. Panel (d) is the intersection of Panels (b) and (c). All regressions include block fixed effects.

Table 3: Treatment effect on basal fertilizer usage
Standardized joint effects across fertilizer types (basal survey)

	(1) Binary fertilizer use consistent with recommendation	(2) Amount of fertilizer applied (kg/ha)	(3) Distance between suggested & applied fertilizer (kg/ha)
Panel A: Full sample			
Treatment	0.232***	0.372***	-0.122***
	(0.034)	(0.067)	(0.036)
Observations	1,308	1,308	1,308
Panel B: Respondents that provided yield data			
Treatment	0.247***	0.363***	-0.129***
	(0.036)	(0.069)	(0.038)
Observations	1,173	1,173	1,173
Panel C: Respondents for which satellite data is available			
Treatment	0.250***	0.456***	-0.131***
	(0.037)	(0.079)	(0.038)
Observations	1,143	1,143	1,143
Panel D: Respondents for which farmer-reported and satellite yields are available			
Treatment	0.260***	0.432***	-0.140***
	(0.037)	(0.077)	(0.039)
Observations	1,108	1,108	1,108

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table reports fertilizer application information for the basal (first) dose, which was emphasized in the advisory messages. Each dependent variable records the standardized joint effect of the indicated outcome across UREA, MOP, DAP, and Zinc, which is an equally-weighted sum across the standardized treatment effects on the outcome for each input type. In column (1), the dependent variable is assigned a value of 1 if the farmer applied the indicated fertilizer type and was advised to, or did not apply the fertilizer type and was advised not to. The variable is coded to 0 if the farmer did not follow recommendations. In column (2), the dependent variable is the amount of fertilizer that was applied during the basal dose, in kilograms per hectare. We use GPS-measured plot area in the denominator if available, and otherwise use reported plot size. Plot area is divided by the portion of a farmer's plot on which they reported applying fertilizer. Column (3) report regressions of the absolute value of the difference between the recommended and applied fertilizer amount in kg/ha. Differences are winsorized at the 99th percentile. Panel (a) includes the full sample. Panel (b) is restricted to respondents that provided 2018 cotton yield data and plot size information. Panel (c) is restricted to respondents for which we have 2018 satellite yield data. Panel (d) is the intersection of Panels (b) and (c). All regressions include block fixed effects.

Table 4: Treatment effect on full season fertilizer use and yields
Midline survey

	(1) Total fertilizer applied (kg/ha)	(2) Distance between suggested & applied fertilizer (kg/ha)	(3) Fertilizer expenditures (Rs 2017)	(4) Cotton yield (kg/ha)
Panel A: Full sample				
Treatment	0.237***	-0.082***	333.316	0.991
	(0.041)	(0.031)	(459.359)	(40.276)
2017 productivity				0.247***
				(0.022)
Observations	1,441	1,441	1,455	1,362
Adjusted R^2			0.131	0.265
Control mean			8,317.397	952.787
Panel B: Respondents that provided yield data				
Treatment	0.239***	-0.076**	231.412	4.697
	(0.043)	(0.033)	(488.200)	(40.560)
2017 productivity				0.248***
				(0.022)
Observations	1,331	1,331	1,331	1,337
Adjusted R^2			0.126	0.268
Control mean			8,478.365	952.293
Panel C: Respondents for which satellite data is available				
Treatment	0.245***	-0.082**	623.243	-5.083
	(0.044)	(0.033)	(474.947)	(41.393)
2017 productivity				0.252***
				(0.022)
Observations	1,294	1,294	1,294	1,287
Adjusted R^2			0.136	0.271
Control mean			8,028.797	961.606
Panel D: Respondents for which farmer-reported and satellite yields are available				
Treatment	0.249***	-0.083**	591.989	-1.132
	(0.045)	(0.033)	(485.275)	(41.684)
2017 productivity				0.252***
				(0.022)
Observations	1,260	1,260	1,260	1,265
Adjusted R^2			0.138	0.274
Control mean			8,150.738	961.211

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Columns (1) - (3) report fertilizer application information across all doses. In columns (1) and (2) the dependent variables record the standardized joint effect of the indicated outcome across UREA, MOP, DAP, and Zinc, which is an equally-weighted sum across the standardized treatment effects on the outcome for each input type. In column (1), the dependent variable is the total amount of fertilizer applied across all doses divided by the average area on which the respondent reported applying fertilizer. Column (2) reports regressions of the absolute value of the difference between the recommended and applied fertilizer amount in kg/ha. Differences are winsorized at the 99th percentile. Column (3) reports total fertilizer expenditures across all doses, winsorized at the 99th percentile. Expenditures are calculated using price data obtained through the Government of India Department of Fertilizers from June 19, 2017. Prices are reported in nominal 2017 Indian rupees. Column (4) reports the treatment effect on agricultural productivity. Productivity is defined as farmer-reported total cotton yield, in kilograms, divided by plot size, in hectares. We use GPS measured plot size when available, and otherwise use the farmer-reported value. We winsorized strictly positive yield values at the 2nd and 98th percentiles. The sample size differs between the columns because the data comes from different surveys and because column (3) does not normalize the dependent variable by plot size. Panel (a) includes the full sample. Panel (b) is restricted to respondents that provided 2018 cotton yield data and plot size information. Panel (c) is restricted to respondents for which we have 2018 satellite yield data. Panel (d) is the intersection of Panels (b) and (c). All regressions include block fixed effects.

Table 5: Treatment effect on basal and full-season fertilizer usage by type
Basal and midline surveys

	(1) UREA	(2) DAP	(3) MOP	(4) Zinc	(5) Standardized joint effect
Panel A: Binary fertilizer decision consistent with recommendation, basal					
Treatment	0.117***	0.0395	0.103***	0.0248	0.232***
	(0.0184)	(0.0251)	(0.0204)	(0.0176)	(0.0341)
Observations	1,308	1,308	1,308	1,308	1,308
Control mean	0.077	0.656	0.116	0.103	
Panel B: Amount of fertilizer applied (kg/ha), basal					
Treatment	11.41***	-4.730	9.628***	0.849***	0.372***
	(2.518)	(6.737)	(2.029)	(0.243)	(0.0669)
Observations	1,308	1,308	1,308	1,308	1,308
Control mean	8.142	87.823	2.212	0.080	
Panel C: Distance between suggested and applied fertilizer (kg/ha), basal					
Treatment	-9.229***	-3.712	-4.749	-0.198	-0.122***
	(1.793)	(3.359)	(3.496)	(0.387)	(0.0358)
Observations	1,308	1,308	1,308	1,308	1,308
Control mean	123.822	64.988	176.627	18.794	
Panel D: Binary fertilizer decision consistent with recommendation, full season					
Treatment	0.0425**	-0.0232	0.115***	0.0390*	0.0962***
	(0.0191)	(0.0242)	(0.0244)	(0.0229)	(0.0219)
Observations	1,455	1,455	1,455	1,455	1,455
Control mean	0.816	0.645	0.351	0.324	
Panel E: Amount of fertilizer applied (kg/ha), full season					
Treatment	26.659***	-0.417	17.571***	0.681	0.237***
	(8.658)	(5.118)	(2.393)	(0.461)	(0.041)
Observations	1,440	1,440	1,438	1,434	1,441
Control mean	173.461	117.360	5.743	1.346	
Panel F: Distance between suggested and applied fertilizer (kg/ha), full season					
Treatment	-16.609**	-2.334	-12.367***	-0.523	-0.082***
	(7.908)	(3.612)	(4.317)	(0.469)	(0.031)
Observations	1,440	1,440	1,438	1,434	1,441
Control mean	289.508	81.829	137.176	14.049	

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table reports fertilizer application information for the basal (first) dose, which was emphasized in the advisory messages, and across all doses. In panels (a) and (d), the dependent variable is assigned a value of 1 if the farmer applied the indicated fertilizer type and was advised to, or did not apply the fertilizer type and was advised not to. The variable is coded to 0 if the farmer did not follow recommendations. In panels (b) and (e), the dependent variable is the amount of fertilizer the respondent reported applying during the basal dose in kilograms divided by the area on which they applied the fertilizer in hectare. Panels (c) and (f) report regressions of the absolute value of the difference between the recommended and applied fertilizer amount in kg/ha. Differences are winsorized at the 99th percentile. In each panel, column (5) reports the average standardized effect across columns (1) - (4), which is an equally-weighted sum across the standardized treatment effects on the outcome for each input type.

Table 6: Predicting farmer-reported yields with satellite data
Endline survey, GPS data, and Sentinel-2 imagery

	(1) NDVI	(2) GCVI	(3) reNDVI	(4) MTCI	(5) LAI
Vegetation index	2.772***	0.372***	4.011***	0.568***	0.909***
	(0.134)	(0.022)	(0.188)	(0.034)	(0.058)
Constant	-0.369***	0.025	-0.278***	-0.706***	0.185***
	(0.059)	(0.052)	(0.052)	(0.092)	(0.047)
Observations	1,291	1,291	1,291	1,291	1,291
Adjusted R^2	0.271	0.227	0.287	0.203	0.215
Placebo adjusted R^2	0.117	0.098	0.134	0.155	0.086
Results with grid FE ($.5km \times .5km$)					
Observations	425	425	425	425	425
p-value	0.000	0.000	0.000	0.005	0.003
Placebo p-value	0.610	0.666	0.268	0.825	0.487

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 reports the results of regressions of farmer reported productivity on satellite vegetation indices (VIs), calculated using Sentinel-2 L2A multispectral satellite imagery. Farmer reported productivity is calculated as farmer-reported total yield in metric tons divided by the area in hectares on which the farmer grew cotton. We use GPS-measured area. The index values were derived by taking the median pixel value observed in the area where the farmer grew cotton. We calculated VI values for 5 satellite images, and then we took the maximum value across the five images for each plot. We removed outlying observations in farmer-reported data by winsoring strictly positive values at the 2nd and 98th percentiles. The placebo adjusted R^2 reports results in which we calculate VIs using placebo plot boundary data and then run the same regressions. The Grid FE results report the results of regressions on yield per hectare on .5km by .5km grid fixed-effects using robust standard errors. Specifically, we report the p-value on the vegetation index value using the actual plot boundary data and the p-value on the vegetation index using the plabebo plot boundary data below the "Results with grid FE" header. The sample size after singletons are dropped, p-value using the actual plot boundary data, and p-value using the placebo plot boundary data are reported.

Table 7: OLS vs 2SLS estimates of the relationship between reNDVI and yields:
Endline survey, GPS data, and Sentinel-2 data

	(1) OLS calibration	(2) OLS calibration Reverse regression	(3) 2SLS calibration	(4) 2SLS calibration Reverse regression
reNDVI	4,049.190***	11964.867***	10129.397***	10162.435***
	(182.469)	(467.029)	(1,176.034)	(1,183.937)
Constant	-198.233***	-2657.420***	-2104.737***	-2114.976***
	(63.307)	(158.338)	(367.617)	(370.591)
Block FE	Yes	Yes	Yes	Yes
Observations	1,197	1,197	1,066	1,066
First-stage F			61.228	125.429
J-test p			0.615	0.615

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7 examines how estimates of the relationship between reNDVI and farmer-reported yields change if we use OLS vs 2SLS to estimate it. All columns exclude farmer-reported yield values equal to 0 or above 3,900 kg/ha to reduce noise. Columns (1) and (3) estimate models directly, whereas (2) and (4) estimate reverse regressions then invert the parameters. We winsorized positive farmer-reported yield values at the 2nd and 98th percentiles. We use total rainfall between June and October 2018 and sowing date as instruments. All regressions include block fixed-effects.

Table 8: Treatment effect on yields (kg/ha): satellite vs. survey data
Endline survey, GPS data, and Sentinel-2 data

	(1) Reported yield and plot size	(2) Reported yield GPS plot size	(3) Satellite yield OLS calibration†	(4) Satellite yield 2SLS calibration
Treatment	-4.850	0.984	-3.008	-7.524
	(38.655)	(42.283)	(21.068)	(52.704)
2017 productivity	Yes	Yes	Yes	Yes
2016 productivity	No	No	Yes	Yes
Block FE	Yes	Yes	Yes	Yes
Observations	1,261	1,261	1,261	1,261
Adjusted R^2	0.248	0.273	0.533	0.280
Control mean	894.276	961.873	952.225	935.513
95% CI:	[-80.687, 70.986]	[-81.969, 83.936]	[-44.341, 38.325]	[-110.922, 95.874]

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

†The standard error in column (3) is likely biased and is included only for comparison.

Table 8 examines how replacing farmer-reported yield measurements with satellite data affects estimates of the treatment effect on agricultural productivity (kg/ha). All columns examine a restricted sample for which none of survey yield data, survey cotton cultivation area, or satellite yield data are missing. Columns (1) and (2) examine farmer-reported total productivity divided by farmer-reported plot size and GPS-measured plot size respectively. We winsorized positive farmer-reported yield values at the 2nd and 98th percentiles. Columns (3) and (4) consider satellite yield measurements obtained by taking the median Red-edge NDVI (Viña and Gitelson, 2005) pixel value contained in each plot for each satellite pass, then taking the maximum reNDVI value across the satellite images and linearly fitting it to farmer-reported yield (using GPS area in the denominator, not farmer-reported plot size). Column (3) considers a satellite prediction model calibrated with OLS, and column (4) examines a model calibrated using 2SLS.

Table 9: Sample sizes needed to detect a 5% change

	(1) Survey yield Survey area	(2) Satellite yield OLS calibration	(3) Satellite yield 2SLS calibration	(4) Satellite yield 2SLS calibration Two-step bootstrap
N, all outcome lags	12,411	3,506	15,524	15,524
N, 1 satellite lag				19,530
N, no outcome lags	12,647			20,932
N, Lee bounds	34,300			

Table 9 displays the sample size needed to detect a 5% change in cotton yields (e.g. kg/ha) with 95% confidence and 90% power. Columns (1) uses farmer-reported productivity divided by farmer-reported plot size. Columns (2) - (4) use Sentinel-2 data. Column (2) uses OLS calibration and columns (3) and (4) use 2SLS calibration. Columns (2) and (3) calibrate based on the relationship between reNDVI and yields on the full sample, whereas column (4) calibrates the model based on a 2SLS regression on each bootstrapped sample. All power calculations were obtained using a bootstrap. Sample sizes for Lee (2009) bounds were obtained based on the lower bound. The top row provides the sample size using all available lags as controls, and subsequent rows show how the sample size changes if controls are excluded.

Appendix A: Figures

Appendix Figure 1: Soil Health Card Developed by PAD

ક્રિષી તરંગ
ખેતી માહિતી સેવા

Soil Health Card		Name of the Lab:				
Farmer's Details		**Soil Test Results**				
Name		**#**	**Parameter**	**Test Value**	**Unit**	**Level**
Village						
Block						
District		**1**	pH	8.26		Acidic
Mobile Number		**2**	EC	0.19	dS/m	Normal
UID		**3**	Nitrogen	50	Kg/Ha	Medium
Soil Sample Information		**4**	Phosphorous	30	Kg/Ha	Medium
MM-YYYY of Collection		**5**	Potash	30	Kg/Ha	Low
Plot Name		**6**	Sulphur	20	PPM	Low
Plot Size		**7**	Zinc	0.80	PPM	Low
Irrigated		**8**	Iron	1.79	PPM	Low

Fertilizer Recommendation											
Secondary & Micro Nutrient Recommendations			**#**	**Crop**	**FYM**	**Fertilizer**					
#	**Fertilizer**	**Quantity**				#	**Fertilizer**	**Basal Fertilizer**	**1 Month After Sowing**	**2 Month After Sowing**	**3 Month After Sowing**
1	**Zinc Sulphate**	**1** Kg/Vigha	1	**Irrigated Cotton**	**10** Ton/ha	**1**	**Urea**	**24** Kg/Vigha	**24** Kg/Vigha	**24** Kg/Vigha	**24** Kg/Vigha
						2	**DAP**	**9** Kg/Vigha	**9** Kg/Vigha		
						3	**Muriate of Potash**	**29** Kg/Vigha			
2	**Gypsum**	**100** Kg/Vigha	2	**Un-Irrigated Cotton**	**10** Ton/ha	1	**Urea**	**10** Kg/Vigha	**10** Kg/Vigha		

Appendix Figure 2: Supplement to Soil Health Card

ક્રિષી તરંગ
ખેતી માહિતી સેવા

Name: Village: Block: UID:

Plot Name:

Summary of Nutrient Levels in Your Plot				
	Macronutrients			Micronutrients
Nutrient Name	**Nitrogen**	**Phosphorus**	**Potash**	**Zinc**
Nutrient Level	Low	High	Medium	Low

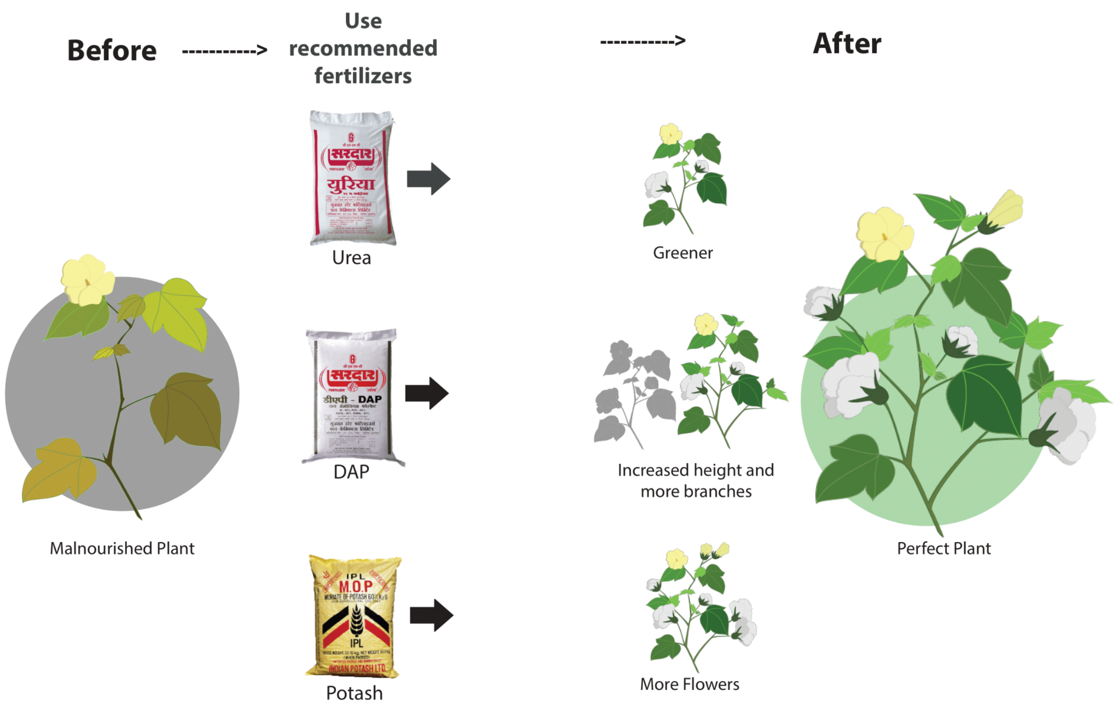
Recommendations of Chemical Fertilizers for Irrigated Cotton					
When to Apply?		 **Urea**	 **DAP**	 **Muriate of Potash (MoP)**	 **Zinc Sulphate**
Basal	At the time of sowing	**24 Kg/Bigha**	**9 Kg/Bigha**	**29 Kg/Bigha**	**1 Kg/Bigha**
Dose 1	1 month after sowing	**24 Kg/Bigha**	**9 Kg/Bigha**	X	X
Dose 2	2 months after sowing	**24 Kg/Bigha**	X	X	X
Dose 3	3 months after sowing	**24 Kg/Bigha**	X	X	X

Application of FYM: Before sowing of Cotton it is advised to apply 1.5 tonn Farm Yard Manure (FYM) per Bigha

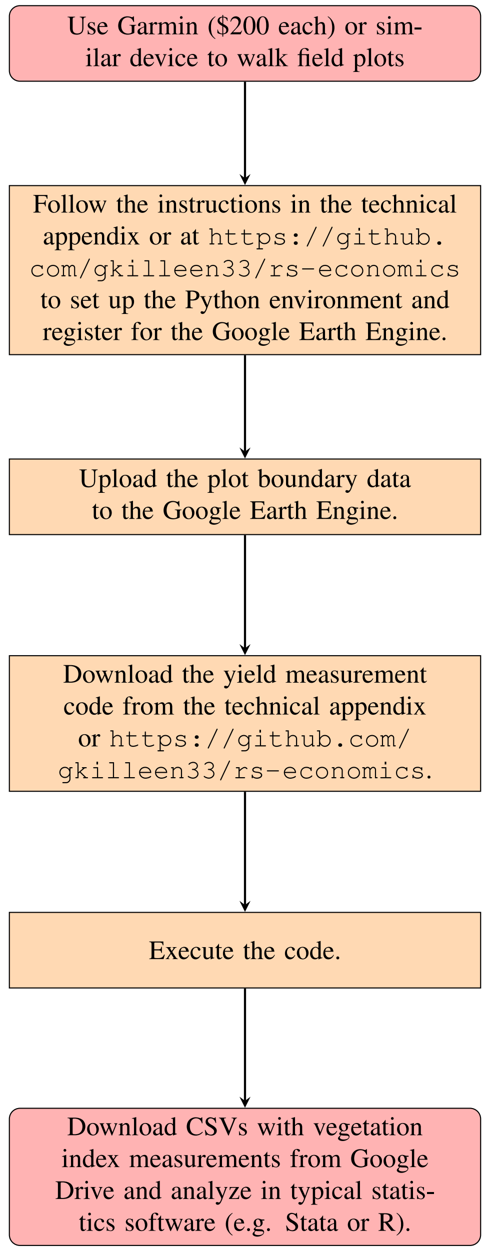
Recommendations of Chemical Fertilizers for Un-Irrigated Cotton		
When to Apply?		**Urea** 
Basal	At the time of sowing	**10 Kg/Bigha**
Dose 1	1 month after sowing	**10 Kg/Bigha**

Application of Gypsum: Before sowing of Un-Irrigated Cotton it is advised to apply 100 kg Gypsum per Bigha

Appendix Figure 3: Booklet



Appendix Figure 4: Process for calculating satellite yield measurements



Appendix B: Tables

Appendix Table 1: Generating recommendations from soil test results

A four-step process was followed to generate customized fertilizer recommendations for each farmer:
1. Before the start of the agricultural season, we collected soil samples from one of the plots owned by each farmer in the sample and tested the soil samples for pH and EC and various macronutrients (nitrogen, phosphorus, potassium) and micronutrients (zinc, sulphur, iron). The soil test results contained the quantity of each nutrient in the soil and the level of each nutrient (low, medium or high).
2. We used nutrient levels to generate nutrient-specific recommendations. For doing so we used a template developed by Junagarh Agricultural University, in which quantities of nutrients are recommended for each of three nutrient levels. The recommended nutrient quantities for every nutrient label are the following:

	Low	Medium	High
Nitrogen (N)	300	240	180
Phosphorus (P)	62.5	50	37.5
Potassium (K)	187.5	150	112.5
Zinc	25	20	15

3. Nutrient levels were then converted to fertilizer recommendations. We focused on three macronutrients (nitrogen, phosphorus and potassium) and one micronutrient (zinc for irrigated cotton and sulphur for unirrigated cotton).
In irrigated plots HYV seeds are used. These seeds require adequate amounts of the three macronutrients selected. Nitrogen and phosphorus are important for crop development and potassium improves water use efficiency, builds crop resilience against certain diseases and improves fibre quality. Application of zinc was also recommended because plants from HYV seeds respond better to macronutrients when micronutrients are available in adequate quantity and most plots in the study area were deficient in this micronutrient.
In non-irrigated or rainfed plots, non-HYV seeds are used. Nitrogen and sulphur were recommended because the nutrient requirements can be met with these two nutrients.
Our fertilizer recommendations were in terms of quantities of UREA, Di-ammonium Phosphate (DAP), Muriate of Potash (MOP), Zinc Sulphate (Zinc) and Gypsum. The table shows the nutrients contained in each fertilizer.

	Nutrient content (%)				
Fertilizer	Nitrogen	Phosphorus	Potassium	Zinc	Sulphur
UREA	46	x	x	x	x
DAP	18	46	x	x	x
MOP	x	x	60	x	x
Zinc Sulphate (Zinc)	x	x	x	36	14
Sulphur	x	x	x	x	100

The nutrient levels in each fertilizer were used to calculate the exact quantity of fertilizer recommended for each plot. Our previous field surveys had shown that all the recommended fertilizers were easily available, reasonably priced and were effective for supplying nutrients to soil.
5. Given that fertilizers are more effective when applied in multiple small doses at various crop stages, total fertilizer recommendations were split into dose-wise recommendations. All doses contained equal quantities of fertilizer. Following are the number of doses in which application of various nutrients is suggested:

	Irrigated Crop		Un-irrigated crop	
	Number of doses	Timing of doses	Number of doses	Timing of doses
Nitrogen	4	- At time of sowing (basal dose) - One month after sowing - Two months after sowing - Three months after sowing	2	- At time of sowing (basal dose) - One month after sowing
Phosphorus	2	- At time of sowing (basal dose) - One month after sowing	0	
Potassium	1	- At time of sowing (basal dose)	0	
Zinc	1	- At time of sowing (basal dose)	0	
Sulphur	1	- At time of sowing (basal dose)	1	- At time of sowing (basal dose)

Fertilizer and nutrient recommendations were generated ‘per unit of area’ to make recommendations farmer friendly. This means that the recommendations were generated for the area unit in which a farmer had reported crop area at baseline. For example, if a farmer had reported land in acres, then customized fertilizer recommendations were made in per-acre terms. Also, since irrigation status of crops is uncertain for farmers in India at the start of the agricultural season, recommendations for both irrigated and unirrigated cotton were generated for each farmer.

Appendix Table 2: Fertilizer knowledge questions

	Question	Correct Answer
1.	Which main nutrients are required by irrigated cotton for growth?	Nitrogen, Phosphorous, and Potash
2.	Which of the following is main nutrient in UREA?	Nitrogen
3.	Which of the following is main nutrient in DAP?	Phosphorous
4.	Which of the following is main nutrient in MOP?	Potash
5.	Which is the best fertilizer for applying nitrogen in the soil?	UREA
6.	Which is the best fertilizer for adding potash in the soil?	MOP
7.	When is zinc recommended to be applied during cotton cultivation for irrigated cotton?	At the time of sowing
8.	When should Urea be applied in the soil for irrigated cotton?	At time of sowing, and 1, 2, and 3 months after sowing
9.	When should Urea be applied in the soil for un-irrigated cotton?	At the time of sowing and one month after sowing
10.	When should DAP be applied in the soil for cotton cultivation?	At the time of sowing, and one month after sowing
11.	When should MoP be applied in the soil for cotton cultivation?	At the time of sowing
12.	What is the benefit of appling UREA to soil?	More green
13.	What is the benefit of applying DAP to soil?	Increased height and more branches
14.	What is the benefit of applying Potash to soil?	More flowers

Appendix Table 3: Survey completion rates

	Control		**Treatment**		**Total**	
	Number	Percent	Number	Percent	Number	Percent
Grew cotton						
Did not sow cotton	45	5.7	38	4.8	83	5.2
Sowed cotton	745	94.3	754	95.2	1499	94.8
Total	790	100.0	792	100.0	1582	100.0
Basal						
Completed basal survey	707	89.3	729	91.9	1436	90.6
Attrited	85	10.7	64	8.1	149	9.4
Total	792	100.0	793	100.0	1585	100.0
Midline						
Completed midline survey	773	97.6	760	95.8	1533	96.7
Attrited	19	2.4	33	4.2	52	3.3
Total	792	100.0	793	100.0	1585	100.0
Endline						
Completed endline survey	736	92.9	729	91.9	1465	92.4
Attrited	56	7.1	64	8.1	120	7.6
Total	792	100.0	793	100.0	1585	100.0
Missing plot map						
Plot mapped	702	88.6	687	86.6	1389	87.6
Plot not mapped	90	11.4	106	13.4	196	12.4
Total	792	100.0	793	100.0	1585	100.0
All surveys and mapping complete						
Did not complete 1+ surveys	168	21.2	175	22.1	343	21.6
Completed all surveys	624	78.8	618	77.9	1242	78.4
Total	792	100.0	793	100.0	1585	100.0

Appendix Table 3 presents the completion rate of each survey and the plot mapping exercise. We determined that a farmer completed a given survey if they grew cotton, the surveyor was able to conduct the survey, and the farmer consented to be surveyed. In two instances, farmers attempted to grow cotton but the crops failed extremely early in the season. We treated these two observations as if the farmers did not grow cotton since they did not apply any inputs.

Appendix Table 4: Attrition

	(1) Basal	(2) Midline	(3) Endline	(4) Farmer-reported yield	(5) Satellite data
Treatment	-0.0441	-0.0154	-0.0562	-0.0721	-0.00108
	(0.0960)	(0.0612)	(0.0880)	(0.127)	(0.109)
Age (baseline)	-0.00151	-0.000234	-0.000528	-0.000344	-0.000683
	(0.00109)	(0.000515)	(0.000907)	(0.00128)	(0.00109)
Literate	-0.0454	-0.0130	-0.0409	-0.0672*	-0.0148
	(0.0339)	(0.0195)	(0.0309)	(0.0406)	(0.0336)
Total cotton land (2017)	-0.00113	0.000971	0.00738	0.00691	0.0139***
	(0.00329)	(0.00233)	(0.00465)	(0.00501)	(0.00517)
Sampled plot size (2017)	0.0115	0.00429	-0.00655	-0.0201*	-0.00407
	(0.00943)	(0.00565)	(0.00877)	(0.0104)	(0.0103)
Irrigation	0.0206	-0.0303	0.0126	-0.0633	0.0643
	(0.0370)	(0.0263)	(0.0317)	(0.0515)	(0.0403)
Strong house	0.0178	-0.00970	-0.00299	-0.0263	-0.0314
	(0.0233)	(0.0129)	(0.0179)	(0.0271)	(0.0226)
Own plough	-0.0522**	-0.00812	-0.0252	-0.0366	-0.0296
	(0.0239)	(0.0126)	(0.0187)	(0.0276)	(0.0230)
Crop insurance	0.0233	0.00474	0.00512	0.00601	0.00284
	(0.0224)	(0.0109)	(0.0178)	(0.0251)	(0.0225)
Children	-0.0168	-0.00120	-0.00785	-0.00556	-0.0141*
	(0.0103)	(0.00370)	(0.00579)	(0.00935)	(0.00758)
$>$ median education	-0.0106	-0.00372	0.00571	0.0308	0.0000631
	(0.0249)	(0.0105)	(0.0191)	(0.0276)	(0.0250)
Soil tested prior to study	-0.00964	0.00143	-0.0162	-0.0305	-0.00812
	(0.0302)	(0.0147)	(0.0261)	(0.0344)	(0.0342)
UREA last season (kg/ha)	0.0000761***	-7.38e-08	-0.0000157**	-0.0000438*	-0.0000265***
	(0.0000110)	(0.00000746)	(0.00000791)	(0.0000234)	(0.0000100)
DAP last season (kg/ha)	-0.0000225	0.0000141	-0.0000130	-0.0000582*	-0.0000283
	(0.0000257)	(0.0000222)	(0.0000223)	(0.0000330)	(0.0000298)
MOP last season (kg/ha)	-0.000118	-0.0000455	-0.000369	-0.000394	-0.000530
	(0.000461)	(0.0000699)	(0.000225)	(0.000419)	(0.000324)
Zinc last season (kg/ha)	-0.00214***	-0.0000666	0.000151	-0.000666	0.00135
	(0.000828)	(0.000358)	(0.000748)	(0.00141)	(0.00138)
Observations	1,585	1,585	1,585	1,585	1,585
Adjusted R^2	0.013	0.011	0.018	0.009	0.048
Control attrition rate	0.107	0.024	0.071	0.148	0.114
p-val of interactions	0.058	0.414	0.533	0.723	0.686

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 4 reports the results of tests for differential attrition. Each regression includes interactions between treatment and each of the other independent variables which are omitted for space. We present the results of an F-test evaluating the joint significance of the interaction terms. A survey is defined as complete if the farmer consented to be interviewed and sowed cotton. Education data was missing for 175 observations. Number of children is missing for 25 observations. Crop insurance is missing for 20 observations. UREA and DAP usage last season are missing for 3 observations, DAP and MOP usage last season are missing for two observations, and sampled plot size is missing for one observation. Missing values of these variables were imputed with the median value of each variable. All regressions include block fixed effects.

Appendix Table 5: Listening rates of treatment calls
Administrative data

	Share of relevant farmers that heard $\geq$ 1 call	Number of relevant farmers
Basal: irrigated (UREA, MOP, and DAP)	0.883	529
Early season: Potash (irrigated)	0.866	529
Dose 2: irrigated (UREA, MOP, and DAP)	0.820	529
Dose 3: irrigated (UREA, MOP, and DAP)	0.788	529
Dose 4: irrigated (UREA, MOP, and DAP)	0.745	529
Early season: Potash (irrigated)	0.866	529
Mid-season: Potash (irrigated)	0.713	529
Early season: Zinc (irrigated)	0.758	529
Mid-season: Zinc (irrigated)	0.437	529
Basal: unirrigated (UREA)	0.760	225
Dose 2: unirrigated (UREA)	0.729	225

Appendix Table 5 reports the share and number of relevant farmers in the treatment group that heard at least 1 customized fertilizer recommendation of the indicated type. A relevant farmer means that they sowed cotton and have an irrigated plot if the advice is for irrigated plots or an unirrigated plot if the recommendation is for unirrigated plots. Customized calls were only sent to farmers in the treatment group. All treatment farmers received the same calls. If the call duration exceeded the point where a recommendation was given, then heard recommendation was assigned a value of 1. The last two rows have a smaller sample size because the majority of farmers reported growing irrigated cotton.

Appendix Table 6: Fertilizer usage by season, dose
Baseline, basal, and midline surveys

	2017 Season					2018 Season				
	Rec. dose	1(applied), control	1(applied), treatment	Dose $\|$(Applied= 1) Control	Dose $\|$(Applied= 1) Treatment	Rec. dose	1(applied), control	1(applied), treatment	Dose $\|$(Applied= 1) Control	Dose $\|$(Applied= 1) Treatment
UREA: Basal dose unirrigated	97.38	0.93*	0.97*	74.27*	93.38*	94.43	0.03	0.08	46.65	63.61
UREA: Basal dose irrigated	131.96	0.99*	0.99*	124.29*	114.06*	134.00	0.08	0.21	108.88	102.69
UREA: Dose 2 unirrigated	98.22	0.79*	0.83*	75.45*	70.86*	94.97	0.61	0.65	97.20	92.63
UREA: Dose 2 irrigated	131.74	0.96*	0.95*	129.20*	139.32*	132.29	0.90	0.94	91.97	92.80
UREA: Dose 3 unirrigated	0.00	0.28*	0.33*	64.00*	72.87*	0.00	0.33	0.42	93.13	97.72
UREA: Dose 3 irrigated	142.23	0.63*	0.63*	117.01*	129.68*	142.43	0.87	0.93	100.76	105.09
UREA: Dose 4 unirrigated	0.00	0.05*	0.01*	69.18*	67.98*	0.00	0.07	0.13	81.93	111.41
UREA: Dose 4 irrigated	143.49	0.21*	0.20*	119.41*	111.53*	143.10	0.80	0.89	105.87	114.76
UREA: Total unirrigated	197.92	0.93*	0.97*	155.96*	150.03*	191.51	0.66	0.75	135.33	138.98
UREA: Total irrigated	546.42	0.99*	0.99*	336.48*	322.79*	554.32	0.85	0.88	249.77	268.21
DAP: Basal dose unirrigated	0.00	0.92*	0.99*	78.47*	78.88*	0.00	0.66	0.72	116.09	100.56
DAP: Basal dose irrigated	57.18	0.91*	0.89*	136.00*	127.03*	57.99	0.69	0.74	128.94	114.92
DAP: Dose 2 unirrigated	0.00	0.20*	0.12*	64.86*	52.42*	0.00	0.22	0.31	93.15	76.05
DAP: Dose 2 irrigated	56.28	0.44*	0.43*	191.81*	232.67*	55.47	0.53	0.59	102.59	98.83
DAP: Total unirrigated	0.00	0.92*	0.99*	92.73*	85.60*	0.00	0.80	0.88	90.53	111.39
DAP: Total irrigated	111.11	0.91*	0.89*	199.98*	187.00*	112.97	0.80	0.81	136.72	132.33
MOP: Basal dose unirrigated	0.00	0.00*	0.01*	*	67.83*	0.00	0.03	0.02	25.48	83.46
MOP: Basal dose irrigated	209.60	0.08*	0.08*	65.69*	72.47*	202.89	0.02	0.14	97.48	89.77
MOP: Total unirrigated	0.00	0.00*	0.01*	*	67.83*	0.00	0.03	0.07	3.09	0.00
MOP: Total irrigated	209.60	0.08*	0.08*	83.18*	80.03*	202.89	0.08	0.25	59.19	91.60
Zinc: Basal dose unirrigated	0.00	0.00*	0.04*	*	16.41*	0.00	0.00	0.00		
Zinc: Basal dose irrigated	18.97	0.07*	0.08*	23.71*	270.35*	19.60	0.01	0.04	12.79	24.09
Zinc: Total unirrigated	0.00	0.00*	0.04*	*	19.17*	0.00	0.01	0.04	11.30	27.13
Zinc: Total irrigated	18.97	0.07*	0.08*	25.99*	32.13*	19.54	0.06	0.11	31.27	22.66

* denotes that the value is estimated and differs from the 2018 construction.

This table reports reports average recommended fertilizer amounts by dose and average actual application by season and treatment status. All recommendations were for the 2018 season, but column (2) differs from column (7) due to differences in irrigation status and extensive margin application decisions. Columns (3) - (4) and (8) - (11) report the share of farmers that reported applying the fertilizer at the dose. Columns (2) and (7) report the average recommended dosage, based on 2018 soil testing data, across farmers that applied any fertilizer. Columns (5) - (6) and (10) - (11) report the average amount applied among those that applied the fertilizer type. In 2018, we collected whether the first dose was applied at sowing (basal application) or later in the season. In 2017, the first dose is treated as the basal dose which inflates application, and the 2018 cotton sowing area is used in calculations. Amount applied by dose is only reported for recommended doses, so total application may exceed the row sum.

Appendix Table 7: Treatment effect on revenue
Midline and endline survey data

	(1) Sold (kg)	(2) Stored (kg)	(3) Sales Revenue (Rs)	(4) Expected Total Revenue (Rs)	(5) Av. Price (Rs/kg)	(6) Revenue minus fertilizer costs (Rs)
Treatment	-6.403	-103.5*	-91.16**	-89.02**	-0.118	8918.8
	(169.4)	(61.38)	(44.50)	(44.65)	(0.248)	(9408.7)
Block FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1336	1342	1245	1341	1239	1306
R^2	0.194	0.008	0.152	0.142	0.060	0.005
Control mean of dependent variable	2470.4	336.1	1678.3	1710.8	2.320	-17000

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 7 reports treatment effect estimates on the amount of cotton sold (column 1), the amount of cotton stored (column 2), sales revenue at the time of the survey (column 3), expected total revenue including stored cotton (column 4), the average price per kg of cotton already sold (column 5), and expected total revenue net of fertilizer costs (column 6). Expected total revenue was calculated from estimates of revenue from existing sales and planned sales of stored cotton. The average price was calculated by normalizing sales revenue by the amount of cotton sold. Revenue net of fertilizer costs was calculated from expected total revenue and fertilizer expenditure data. Fertilizer expenditure data is from the midline survey, resulting in a smaller sample. All other parameters were directly measured in surveys. All regressions include block fixed effects and all outcomes are winsorized at the 99th percentile.

Appendix Table 8: Satellite data availability
GPS and Sentinel-2 data

	Control		Treatment		Total	
	Number	Percent	Number	Percent	Number	Percent
Plot mapping						
Not mapped	85	11.3	105	13.8	190	12.5
Mapped	670	88.7	656	86.2	1,326	87.5
Sentinel-2: 2016						
No data	2	0.3	1	0.2	3	0.2
Data	668	99.7	655	99.8	1,323	99.8
Sentinel-2: 2017						
No data	2	0.3	1	0.2	3	0.2
Data	668	99.7	655	99.8	1,323	99.8
Sentinel-2: 2018						
Data	670	100.0	656	100.0	1,326	100.0
Sentinel-2: no missing data						
Missing data	2	0.3	1	0.2	3	0.2
No missing data	668	99.7	655	99.8	1,323	99.8

Appendix Table 8 reports the availability of cloud-free Sentinel-2 satellite imagery by treatment status. Data is non-missing for a plot in a year if at least one cloud free image was available. There is no data available for 3 plots in 2016 and 2017 because they fall outside of the Sentinel-2 swath that images the majority of the sample.

Appendix Table 9: Vegetation indices

Index	Equation	Equation in Sentinel-2 bands	Source
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - Red}{NIR + Red}$	$\frac{B8 - B4}{B8 + B4}$	Haas et al. (1974)
Green Chlorophyll Vegetation Index (GCVI)	$\left(\frac{NIR}{GREEN}\right) - 1$	$\left(\frac{B8}{B3}\right) - 1$	Gitelson et al. (2003)
Red-edge NDVI (reNDVI)	$\frac{NIR - RE}{NIR + RE}$	$\frac{B8 - B5}{B8 + B5}$	Viña and Gitelson (2005)
MERIS Terrestrial Chlorophyll Index (MTCI)	$\frac{NIR - RE}{RE - Red}$	$\frac{B8 - B5}{B5 - B4}$	Dash and Curran (2004)
Leaf Area Index (LAI)	Neural net	Weiss and Baret (2016)	Weiss and Baret (1999)

Sentinel-2 bands 3, 4, and 8 have a 10 meter by 10 meter resolution, but band 5 has a 20 meter by 20 meter resolution. Hence, we resampled band 5 using a bilinear algorithm before deriving the Red-edge NDVI and MTCI. We resampled all Sentinel-2 bands to 10 meters before calculating LAI.

Appendix Table 10: Satellite-measured vs farmer-reported yields
Endline survey, GPS data, and Sentinel-2 imagery

	(1) NDVI	(2) GCVI	(3) reNDVI	(4) MTCI	(5) LAI
Panel A: Plots above the median size					
Yield (metric tons/hectare)	0.103***	0.573***	0.073***	0.334***	0.230***
	(0.006)	(0.045)	(0.004)	(0.026)	(0.018)
Constant	0.356***	1.779***	0.221***	2.530***	0.559***
	(0.008)	(0.048)	(0.005)	(0.033)	(0.019)
Observations	647	647	647	647	647
Adjusted R^2	0.321	0.252	0.332	0.221	0.244
Panel B: Plots below the median size					
Yield (metric tons/hectare)	0.096***	0.664***	0.072***	0.390***	0.251***
	(0.008)	(0.059)	(0.005)	(0.032)	(0.023)
Constant	0.416***	2.088***	0.259***	2.665***	0.698***
	(0.009)	(0.059)	(0.006)	(0.038)	(0.023)
Observations	644	644	644	644	644
Adjusted R^2	0.250	0.231	0.270	0.204	0.214
p-value: R^2 Panel A $\leq R^2$ Panel B	0.062	0.330	0.085	0.344	0.261

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 10 reports the results of regressions of satellite vegetation indices (VIs), calculated using Sentinel-2 L2A multispectral satellite imagery, on farmer reported productivity. We calculated the median VI value observed in each plot for 5 Sentinel-2 images from 2018 and then took the maximum value across the 5 satellite passes for each plot and each VI. We removed outlying observations in farmer-reported data by winsoring strictly positive values at the 2nd and 98th percentiles. We report the probability that the increase in adjusted R^2 from panel (a) to panel (b) is due to random chance in each column. These p-values were obtained using randomization inference.

Appendix Table 11: The effect of irrigation on yields (kilograms/hectare)
Survey data, GPS data, and Sentinel-2 imagery

	(1) Survey yield Survey area	(2) Survey yield GPS area	(3) Satellite yield
Irrigation	700.830***	818.289***	122.891***
	(56.431)	(71.121)	(32.876)
2018	-862.495***	-919.476***	-1304.268***
	(53.904)	(63.794)	(32.789)
Irrigation x 2018	11.374	-47.413	379.344***
	(62.501)	(73.845)	(36.388)
Constant	1,117.736***	1,186.079***	1,813.094***
	(49.572)	(61.991)	(30.871)
Observations	2,522	2,522	2,522
Clusters	1,261	1,261	1,261
Adjusted R^2	0.249	0.221	0.583

Clustered standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table 11 compares regressions of farmer-reported yields and satellite-measured yields on year dummies, irrigation status, and an interaction between irrigation status and year. In column (1), the dependent variable is farmer-reported total cotton yield, in kilograms, divided by cultivation area, in hectares. In columns (2) and (3), the dependent variable is satellite-measured productivity. We calculated satellite-measured yield by calculating, from Sentinel-2 imagery, the median Red-edge NDVI (reNDVI) pixel value observed in each plot for each satellite pass. We then linearly transformed the reNDVI values to productivity using a linear regression on farmer-reported yield divided by GPS-measured plot size. Column (2) uses Sentinel-2 observations from October 28, 2017 and October 28, 2018 only. Column (3) includes all 2017 and 2018 Sentinel-2 passes that we analyze in this paper. Irrigation is defined as 1 if the farmer reported irrigating their plot using underground water, a nearby water source/dam, or a canal during the baseline survey. In 491 instances, farmers indicated that they had an irrigation system fed by rainfall. We coded irrigation to 0 in these cases since the variable is intended to capture whether the respondent has an irrigation system independent of rainfall. We removed outlying observations in farmer-reported yield data by winsoring strictly positive values at the 2nd and 98th percentiles. The sample is restricted to observations for which none of the dependent variables are missing.

8 Technical Appendix

8.1 Satellite yield data and methodology

We construct satellite yield estimates using GPS plot boundary data collected concurrently with the endline survey and multispectral satellite imagery from the European Space Agency's Sentinel-2 mission. In omitted analysis, we also examined PlanetScope imagery offered by the private company Planet Labs. However, PlanetScope did not offer improved yield measurements and has high access barriers, so we excluded the data source from our analysis. Several other recent studies, such as Lobell et al. (2019), have similarly found that satellite constellations such as PlanetScope – which have superior spatial resolution but inferior spectral resolution compared to Sentinel-2 – do not improve yield estimates.

Surveyors collected GPS boundary data by walking the boundaries of plots with Garmin eTrex 30x GPS devices. The research team manually verified the accuracy of each plot boundary using high-resolution satellite imagery. The mapping exercise was completed at the same time as the endline survey, but a separate survey team conducted the plot mapping. The plot mapping and survey teams often met separately with farmers, resulting in a different set of attriters for the endline survey and mapping exercise.

Surveyors were able to collect boundary data for 1,389 of the 1,585 plots visited during the baseline survey. We excluded 63 plots on which cotton was not cultivated from the sample, resulting in 1,326 plot maps of cotton plots. In 23% of the sample, farmers only grew cotton on part of their plot. In these instances, surveyors mapped the plot twice: they first walked the boundary of the entire plot, then they mapped only the portion where cotton was cultivated. We use the cotton cultivation area in place of full plot boundary data in these cases and refer to this area as "plot size" and "plot area." Results are similar if we use the full plot area, although the fit between farmer-reported and satellite-measured yields decreases slightly. The average cotton cultivation area was 1.57 hectares, the standard deviation in cotton cultivation area was 1.12 hectares, and 9% of farmers in the sample sowed cotton on an area of 0.5 hectares or smaller.

We use satellite data from the Sentinel-2 constellation operated by the European Space Agency's Copernicus program. The constellation consists of two identical satellites that operate in an identical orbit, but are placed 180 degrees apart. The satellite constellation collects imagery of most of the planet at least once every 5 days. The spatial resolution of Sentinel-2 images varies for different spectral wavelengths. We use red, green, blue, and near-infrared bands with a spatial resolution of 10 meters by 10 meters (meaning that each pixel in an image corresponds to a 10 meter by 10 meter area on Earth). We also examine red-edge infrared and short-wave infrared bands that have a 20 meter by 20 meter spatial resolution.[22] Sentinel-2 imagery is freely accessible through the Copernicus Open Access Hub, the Google Earth Engine, and a variety of other sources.

We searched for images with less than 20% cloud cover between August 15th and November 15th in 2016, 2017, and 2018. We chose this range of time because it falls between the early flowering period and the start of harvesting in the study sample. Zhao et al. (2007) find that multispectral imagery obtained during this portion of the phonological cycle can accurately predict cotton productivity. Cloud cover near 20% is very high, and may produce invalid results even in cloud-free areas because factors such as cloud shadows interfere with image quality. However, cloud cover is calculated per tile – the 100x100km images that Sentinel-2 data is distributed in – by the Copernicus Open Access Hub from which we downloaded the data, not across the entire AOI. Part of the study sample is contained in a tile that also includes a coastal region with persistent cloud cover. Hence, setting stricter parameters for cloud cover caused data from this portion of the sample to get dropped, even though there were frequently no clouds over the sample plots in this portion of the data.

To solve this problem, we identified dates for which all Sentinel-2 tiles intersecting the sample region had less than 20% cloud cover, and then downloaded the imagery for these dates. We then manually examined the images, and selected dates for which cloud cover was very low. We also excluded several dates that only covered a portion of the sample, particularly in 2017, due to the high computational costs of processing the imagery. We do not have objective measurements of the cloud cover over the sample plots on the selected dates, but we estimate that clouds cover less than 5% of the area of the sample plots for each selected date. This methodology is imperfect since it depends partially on subjective judgment. However, we note that the Google Earth Engine scripts that we include in the GitHub repository accompanying this study (`https://github.com/gkilleen33/rs-economics`) calculate the rate of cloud cover over the area of interest that is uploaded and therefore avoid this problem.

[22]Sentinel-2 also has 60m resolution bands for detecting aerosol, water vapor, and cirrus that are used in atmospheric correction and cloud masking algorithms but not in analysis.

Sentinel-2 images from 2016-10-28, 2017-10-08, 2017-10-28, 2017-11-02, 2017-11-07, 2018-09-28, 2018-10-03, 2018-10-18, 2018-10-23, 2018-10-28, 2018-11-04, and 2018-11-07 were selected using this procedure. Later analysis indicated that imagery obtained prior to October 15th was substantially less predictive of yield. Hence, we dropped the data from 2017-10-08, 2018-09-28, and 2018-10-03 and were left with one image from 2016, three images from 2017, and five images from 2018. None of the images covers the entire sample. A small number of plots fall on the border of a Sentinel-2 swath, meaning that they are imaged several days apart. As a result, the image from 2018-11-04 only covers about 45% of the sample, whereas each of the other images cover more than 99% of the sample. Fewer than 10 plots fall outside of the second Sentinel-2 swath, with the exact number varying across satellite passes. Appendix Table 8 reports the number of treatment and control plots that were mapped, and the number of plots that we were able to obtain cloud-free Sentinel-2 data for in each year. At least one cloud free image was available for each mapped cotton plot in 2018, and at least one cloud free image was available in 2016, 2017, and 2018 for all but three mapped cotton plots.

We downloaded data from the Copernicus program in the Level-1C (L1C) format which is a top-of-atmosphere (TOA) reflectance product. We processed the imagery using version 2.8 of a third party plugin called "Sen2Cor" that generates a bottom-of-atmosphere (BOA) corrected surface reflectance (L2A) product.[23] One of the primary purposes of the program is to remove the effects of the atmosphere on reflectance values, meaning images produced are more accurate representations of the Earth's surface. The program also performs terrain corrections, cirrus corrections, reduces heterogeneity between different tiles, and generates a scene classification (Mueller-Wilm, 2019). We masked out pixels classified as saturated or defective, cloud shadows, a medium cloud probability, or a high cloud probability. In other words, these pixels were coded to missing and excluded from calculations. We did not use the Google Earth Engine for our analysis because the platform did not offer Sentinel-2 L2A imagery covering the sample region and dates of imagery when we conducted the analysis, and we found that the cloud classification produced by "Sen2Cor" was generally accurate.

Using the processed imagery, we constructed 6 vegetation indices (VIs) to measure agricultural productivity. Specifically, we calculated the Normalized Difference Vegetation Index (NDVI), Green Chlorophyll Vegetation Index (GCVI), Red-edge NDVI (reNDVI), MERIS Terrestrial Chlorophyll Index (MTCI), and Leaf Area Index (LAI) for each date. NDVI, GCVI, reNDVI, and MTCI are simple combinations of spectral bands and exploit the fact that chlorophyll absorbs visible red light but reflects infrared light to assess crop health. We selected these VIs based on their effectiveness at measuring maize yields in Burke and Lobell (2017) and Lobell et al. (2020). We estimated LAI using a tool included in the Sentinel Application Toolbox (SNAP) version 7.0.0. The software uses a neural network to estimate LAI. Details of the algorithm are presented in Weiss and Baret (2016). We chose to consider LAI based on the strength of the metric at predicting smallholder productivity in Lambert et al. (2018). Appendix Table 9 defines each of the VIs that we consider and provides a citation for each index source.

Constructing the VIs results in a set of 36 single band images: one for each VI on each date. These images are essentially large arrays where the value of a cell, a pixel in the image, represents the VI value at a point on Earth, and the location of the cell plus metadata map each cell to a geographical location. For each plot and each image, we take the median value of all pixels whose centroid is contained in the plot boundary polygon, excluding any pixels that we masked out in the earlier step. This results in a data set (in this case a CSV) containing a real number for each plot, VI, and date that can be analyzed in standard statistical software.

There are multiple approaches to convert a time-series of VI values into a yield measurement for a plot. For instance, Lobell et al. (2020) regress yield on each of the VI values captured during a season. We adopt the approach used in Lambert et al. (2018) and take the maximum of the VI values observed in each plot during each season. We do so because the approach is more robust to differences in sowing time, allows us to use imagery from both Sentinel-2 swaths covering the sample, and to avoid over-fitting to farmer-reported productivity data that may be inaccurate. This approach also minimizes attrition since only a single cloud-free image is needed per season.

We next assess the effectiveness of satellite VIs at measuring agricultural productivity in this setting by comparing the maximum VI values from 2018 to farmer-reported productivity. For each vegetation index (VI), we estimate the OLS regression

$$VI_i = \alpha + Yield_i + \epsilon_i$$

where $VI_i = \max\{VI_{i,20181018}, VI_{i,20181023}, VI_{i,20181028}, VI_{i,20181104}, VI_{i,20181107} : VIi, d \in R\}$ is the

[23]The ESA recently started providing official L2A products processed using Sen2Cor. However, official L2A images covering our study region and time periods are not available.

maximum vegetation index value observed in plot i and $Yield_i$ is farmer-reported yield in metric tons per hectare. Metric tons are used instead of kilograms to reduce the number of leading zeros in regression results.

8.2 Satellite yield validation

Table 6 presents the results of regressions between each of the VIs and farmer-reported productivity. We find strong evidence supporting the accuracy of satellite yield measurements which is detailed in Section 4.1. The results of two additional tests aimed at assessing the quality of satellite yield measurements are presented in Appendix Table 10 and Appendix Table 11.

In Appendix Table 10, we examine the relationship between each of the VIs and farmer-reported productivity separately among small and large plots. Burke and Lobell (2017) note that the fit between VIs and farmer-reported yield should be better on large plots than small plots if satellite yields are valid. Some sources of noise in farmer-reported data, such as rounding, should decrease as plot size increases because the error is spread over a larger denominator. Larger plots also contain more satellite pixels and the ratio of pixels that are fully contained in the plot to pixels that partially overlap with areas outside of the plot is higher, so measurement error may also be decreasing in satellite yield estimates. Hence, we would expect the R^2 between the VIs and farmer-reported productivity to increase with plot size if satellites are returning an accurate estimate of productivity.

We test this hypothesis by splitting the sample into plots greater than the median size in panel (A) and plots that are less than or equal to the median size in panel (B). The adjusted R^2 associated with each VI is higher in panel (A) than panel (B). For instance, the adjusted R^2 decreases by about 0.062 from panel (A) to panel (B) in the case of reNDVI, which is the best performing VI. We estimate the likelihood that each of the differences is due to random chance using randomization inference. We run 10,000 simulations in which we randomly split the sample into groups of 647 and 644 plots, then examine what fraction of the time

$$[R^2(Group_1) - R^2(Group_2)] > [R^2(Panel_A) - R^2(Panel_B)]$$

for each VI. We report the results of this test at the bottom of Appendix Table 10. The probability that the increase in adjusted R^2 from panel (B) to panel (A) is due to random chance in the case of reNDVI is less than 0.1. Hence, we find relatively strong evidence that the fit between reNDVI and farmer-reported productivity increases with plot size, supporting the validity of satellite measurements of small-holder cotton productivity in this sample.

Appendix Table 11 regresses farmer-reported yield and satellite-measured yield on baseline irrigation, year, and an interaction between irrigation and year. Figure 5 demonstrates that rainfall was well below average during the 2018 growing season, but typical during the 2017 season. Hence, we expect that the productivity difference between plots with irrigation and plots without irrigation should be larger in 2018 than 2017. The interaction between irrigation and year is not significant in the case of farmer-reported productivity in Columns (1) and (2). However, the interaction between treatment and year is positive and statistically significant using satellite yield measurements in Column (3). We measure yield by linearly fitting reNDVI values to farmer-reported productivity. We interpret these results as suggestive evidence that satellite data produces a more precise estimate of true productivity than survey data.

Overall, we find strong evidence that satellite imagery produces accurate measurements of small-holder cotton productivity in Gujarat, India. Red-edge NDVI is the vegetation index most strongly correlated with farmer-reported productivity in this sample. As a result, this index is the basis of satellite yield measurements presented elsewhere in this paper.

As discussed in detail in the main paper, we construct satellite yield estimates from the raw reNDVI values using two approaches. First, we proceed by fitting an OLS regression of farmer-reported productivity on reNDVI and then calculating a linear prediction. Second, we use two-stage least squared to instrument for reNDVI using sow date and total rainfall from June-October, controlling for block fixed-effects. Farmer-reported productivity (in kg/ha) is defined as farmer-reported total yield divided by GPS-measured plot size in these regressions. We use GPS measured plot size since previous research has shown that farmer-reported plot size can introduce bias into productivity measurements.

8.3 Power calculations

Table 9 presents a comparison of power calculations conducted using farmer-reported yield data to power calculations using satellite-measured yields based on the data from this study. This section details the methodology used to construct the power calculations.

Each column presents the estimated total sample size needed to detect a 5% change in yields with 95% confidence and 90% power. The power calculations were also constructed using a standard non-parametric bootstrap. We conducted the bootstrap in Python to allow for better parallelization. For all estimates other than those involving Lee Bounds, our approach is:

1. Create a Pandas dataframe containing all control observations that are not missing farmer-reported plot area data, farmer-reported yield data from 2017 or 2018, or satellite vegetation index measurements for either year.
2. For each sample size N on a grid with a step size of 100:
   - Construct 1,000 bootstrapped samples from the dataframe of size N by sampling with replacement.
   - Assign a randomized treatment variable, T_i, stratifying by block.
   - For each outcome measurement y_i, impose the alternative hypothesis by calculating $y_i' = y_i + .05 * T_i * y_i$.
   - Estimate the regression of interest using y_i' on the bootstrapped sample. The test rejects if $\hat{\beta} > 0$ and the p-value is below .05.
   - Calculate power as the share of the time the test rejects.
3. Find the minimum N over the grid for which power is at least 0.9.
4. Finally, to account for attrition we estimate that the required sample size is N/a where a is 1 minus the attrition rate for the data source. This is the value reported in Table 9.

Note that since we account for attrition in the last step, the estimated sample sizes need not be in multiples of 100 even though the grid is.

To estimate the required sample size giving Lee bounds, we follow a slightly different approach:

1. Include all observations in a Pandas dataframe, including those that attrited.
2. For each sample size N on a grid with a step size of 100:
   - Construct 1,000 bootstrapped samples from the dataframe of size N by sampling with replacement, stratified by treatment assignment and block.
   - For each bootstrapped sample, drop observations missing outcome data.
   - For each outcome measurement y_i, impose the alternative hypothesis by calculating $y_i' = y_i + .05 * T_i * y_i$.
   - Using the actual treatment assignment, T_i, trim the data in order to obtain the lower Lee (2009) bound.
   - Estimate the regression of interest using y_i' on the bootstrapped sample. The test rejects if $\hat{\beta} > 0$ and the p-value is below .05.
   - Calculate power as the share of the time the test rejects.
3. Find the minimum N over the grid for which power is at least 0.9.
4. This is the value reported in Table 9.

This method differs slightly in that we use actual treatment data, rather than limiting analysis to the control group. We believe this is unlikely to affect results since we find no evidence of a treatment effect on yields. The advantage is that we are able to capture the true distribution of differential attrition in the data.

For robustness, we also performed power calculations analytically (which are not reported in this version of the paper), and we found similar results.

The first column uses farmer-reported yield measurements normalized by farmer-reported plot size. The second column uses satellite data calibrated with OLS. The third and fourth columns use satellite data calibrated via two-stage least squares. In columns 2 and 3, we first calculate yield predictions and then perform the bootstraps. In column 4, we calibrate the yield predictions after drawing each bootstrapped sample.

The first row reports estimated sample sizes needed to detect a 5% effect size 90% of the time including all available lagged outcomes as controls. The second row controls only for one year of past satellite data, rather than two, which we only report for column 4 for computational reasons. Row 3 reports sample sizes without controlling for any lagged outcomes, which we again only report for one satellite outcome for computational reasons. Finally, the fourth row reports sample sizes needed to detect an effect size after taking the lower Lee (2009) bound, which is only for farmer-reported data since differential attrition is not possible in the satellite data.

8.4 Instructions for calculating satellite yield estimates

This section provides step-by-step instructions for calculating satellite yield measurements. These instructions use Sentinel-2 L2A data on the Google Earth Engine. This data is available globally beginning in December 2018. Sentinel-2 L1C data is available on the Google Earth Engine for earlier dates, and L2A data can be generated using different tools. However, this documentation does not cover that use case.

Excluding the time taken to collect plot boundary data, these instructions should take less than two hours to complete.

1. Collect plot boundary data for each plot. The boundary data should be in a format such as an ESRI Shapefile or GeoJSON, where each plot is a separate polygon. In addition, each plot should have a unique ID associated with it. This can be added using free software such as QGIS by editing the attribute table if a unique ID is not already present. This will later be used to merge the vegetation index values with other data sources.

   Plot boundary data is most accurate if it is collected by having surveyors walk plot boundaries with GPS devices. Garmin eTrex 30x units were used in this study. Smart phones or tablets may be adequate in some conditions, although they often have worse signal reception and may have lower accuracy.

   If plots cannot be mapped but a single GPS point is available, one can trace the plot boundary using high-resolution satellite imagery and GIS software such as ArcGIS or QGIS.

2. Register for the Google Earth Engine at `https://code.earthengine.google.com/`. Registration will be tied to your Google account. If an academic email address is used, the service is free.

3. Install Python 3 and the Google Earth Engine API. The Conda package manager is suggested for creating the environment. Instructions are available at `https://developers.google.com/earth-engine/python_install-conda`.

4. Upload the plot boundary data to the Google Earth Engine through the Earth Engine console at `https://code.earthengine.google.com/`. Instructions for uploading the plot boundary data are available at `https://developers.google.com/earth-engine/importing`. Make sure that the plot boundary data has an attribute feature that uniquely identifies each polygon. The script will produce panel data in long format, and so an ID variable is essential to analyze the outputted data.

5. Create an ESRI Shapefile or similar file that defines the Area of Interest (AOI) for your project. The AOI should consist of 1 or more polygons that contain all of the plots in your sample. Instructions for creating such a file in QGIS are available at `https://docs.qgis.org/2.8/en/docs/training_manual/create_vector_data/create_new_vector.html`.

6. Upload the AOI data to the Google Earth Engine following the same steps used for the plot boundary data.

7. Create a folder on your Google Drive account that will store the outputted vegetation index values for each plot. This document will be a csv.

8. Open a text document and paste in the code block included at the bottom of these instructions, beginning with `# -*- coding: utf-8 -*-` and ending with `task.start()` followed by an empty line. Save the file with the extension `.py`. If you close the text editor and wish to edit this file in the future, you may need to right click on the document and select open with, then select a text editor or open the file through a text editor. These instructions assume that the file is called `calculate_vis.py` and refer to the file using that name. However, you may title it whatever you wish, as long as it ends with `.py`.

9. Nine values must be inserted into the code for it to work. All values should be inserted in the clearly labeled section for user inputs, and no other sections should be modified.

   - **Plot boundary data**: Insert the Google Earth Engine asset ID for the plot boundary data, wrapped in single or double quotes, between the parentheses in the line `plot_boundaries = ee.FeatureCollection()` which is line number 10. Instructions for finding the Google Earth Engine asset ID are available at `https://developers.google.com/earth-engine/asset_manager#importing-assets-to-your-script`.

- **AOI**: Insert the Google Earth Engine asset ID for the AOI, wrapped in single or double quotes, between the parentheses in line number 11, `aoi = ee.FeatureCollection()`.
- **Start date:** Enter the earliest date from which images should be obtained. Images may not be downloaded from the start date since Sentinel-2 does not image each area of the world every day and many images have cloud cover. This is the earliest date that images will be considered. The date should be entered in the format YYYY-MM-DD between the parentheses in line 15.
- **End date:** Enter the latest date for which images should be obtained in line 16.
- **Maximum cloud coverage:** In line 19, enter the maximum cloud coverage for which images will be downloaded, in percentage terms (with no percent sign). The value 10 in the code `max_cloud_cover = 10` should be replaced with the value that you choose. Cloud cover is calculated across the AOI that you upload, and images with higher cloud coverage will not be considered. The script will also mask out clouds, meaning if a plot falls below a cloud in an image the data will be coded to missing. Values should generally be kept as low as possible for this parameter because even moderate cloud cover can dramatically reduce the quality of results.
- **Minimum AOI coverage:** This parameter specifies the minimum percent of the AOI that should be covered by satellite images obtained on a certain date for the imagery to be downloaded. Since Sentinel-2 covers different areas of the globe on different days, the satellite may only pass over a portion of the sample on a date. Users may want to exclude dates that cover only part of the sample in favor of those that cover the full sample. Enter a number between 0 and 1 in line 22. The default value is 0.9.
- **Reducer:** A given plot typically contains multiple satellite pixels. This parameter determines how to aggregate multiple pixel values into a single measurement per plot for each satellite pass. The value should be entered in line 27 and be contained in single or double quotes. The options are mean, median, min, max, mode, and sd, where sd is standard deviation. This paper uses median to reduce the influence of outliers, but we also considered mean. If you would like to test multiple options, you may run multiple versions of the code.
- **Output folder:** Enter the name of the folder that you created on Google Drive to store the outputs in line 31. The value must be contained within single or double quotes.
- **Output file:** Enter the name you would like the output file to have. Do not include a file extension. The value should be inserted in line 32 and must be contained in single or double quotes.

10. Save `calculate_vis.py` with the updated parameters.
11. Open a command line and change the working directory to the folder containing `calculate_vis.py`.
12. Make sure that you have the Python environment containing the Google Earth Engine API activated. If you are confused by this step, refer to this documentation: https://developers.google.com/earth-engine/python_install-conda
13. In the command line, enter `python calculate_vis.py` and hit enter.
14. The Google Earth Engine will now calculate the vegetation index values for several different VIs for each satellite image meeting the coverage, cloud cover, and date parameters. The output will be panel data in a CSV format saved to the Google Drive folder selected once the process is complete. This can be imported into any statistical software.
15. The process of translating VI values to yield estimates varies. In this paper, we first take the maximum Red-edge Normalized Difference Vegetation Index value for each plot across a season, then linearly fit this value to farmer-reported cotton yields. As discussed in the paper, we recommend using a 2SLS using an instrument such as rainfall to calibrate this relationship. We suggest looking at several of the methods used in papers that we cite, and then selecting the method that seems the most appropriate and works the best in your sample.

```
1 # -*- coding: utf-8 -*-
2
3 import ee
4
```

```
ee.Initialize()

###############################################################
# ENTER USER INPUTS HERE
###############################################################
plot_boundaries = ee.FeatureCollection() # Upload plot boundary data (e.g. using the Google Earth Engine console) and insert the asset ID here, in single or double quotes
aoi = ee.FeatureCollection() # Upload AOI polygon (e.g. using the Google Earth Engine console) and insert the asset ID here, in single or double quotes
# The AOI can contain multiple polygons, but should be relatively simple. A bounding box or convex hull around all plot boundaries is appropriate

# Start and end dates for image search (YYYY-MM-DD)
begin = ee.Date() # E.g. ’2019-08-01’
end = ee.Date() # E.g. ’2019-12-15’

# Set max cloud cover (%)
max_cloud_cover = 10  # This is a default value. This value should be kept relatively small since high cloud cover can bias results.

# Minimum AOI coverage
min_aoi_coverage = 0.9 # This is the minimum area coverage of low cloud satellite tiles (< max_cloud_cover) calculated against AOI
# The percent of plots with data may be higher or lower since they may not be uniformly distributed in the AOI or plots may be beneath clouds

# Reduction method
# One of ’mean’, ’median’, ’min’, ’max’, ’mode’, ’sd’
reducer = ’mean’  # Each plot has multiple pixel values in it. This specifies how a single value should be extracted for each plot on each date.
# For example, if ’mean’ is selected the average pixel value within a plot is calculated and passed to the CSV

# Export information (to Google Drive)
output_folder = ’EXAMPLE_FOLDER’  # Folder name to save outputs in Google drive. The folder should be created before running the script.
output_file = ’EXAMPLE_FILE_NAME’ # Output file name

##############################################################
# END USER INPUTS
##############################################################

if reducer == ’median’:
    ee_reducer = ee.Reducer.median()
elif reducer == ’mean’:
    ee_reducer = ee.Reducer.mean()
elif reducer == ’min’:
    ee_reducer = ee.Reducer.min()
elif reducer == ’max’:
    ee_reducer = ee.Reducer.max()
elif reducer == ’sd’:
    ee_reducer = ee.Reducer.stdDev()
else:
    raise Exception(’Please select a valid reduction method.’)


l2a = ee.ImageCollection("COPERNICUS/S2_SR")

# Filter the Sentinel-2 data based on AOI and cloud cover
filtered = l2a.filterDate(begin, end).filterBounds(aoi).filter(ee.Filter.lte(’CLOUDY_PIXEL_PERCENTAGE’, 15))

# Create a separate image collection by day
number_of_days = end.difference(begin, ’day’)
def calculateDays(day):
    return begin.advance(day,’day’)

list_of_days = ee.List.sequence(0, number_of_days.subtract(1)).map(calculateDays)
```

```python

"""
NOTE
Mosaicing defaults to EPSG:4326 with 1 degree by 1 degree scale by default if there are competing
inputs. This doesn't appear to change the native resolution in this case, but reprojection
will need to occur here to standardize all band projections if the issue emerges.
"""

def calc_footprint(image, list):
    # Cast
    image = ee.Image(image)
    list = ee.List(list)
    tile_footprint = ee.Algorithms.GeometryConstructors.Polygon(
        ee.Geometry( image.get('system:footprint') ).coordinates()
        )
    return list.add(ee.Feature(tile_footprint))


def create_mosaics(date, newlist):
    # Cast values
    date = ee.Date(date)
    newlist = ee.List(newlist)

    # Filter collection between date and the next day
    filtered_day = filtered.filterDate(date, date.advance(1,'day'))

    # Create a variable recording the footprint of the mosaic
    footprint_collection = ee.FeatureCollection(ee.List(filtered_day.iterate(calc_footprint, ee.List([]))))
    footprint = ee.Feature(footprint_collection.union(100).first())

    # Generate the mosaic
    image = ee.Image(filtered_day.mosaic()).set({'Date': date, 'footprint': footprint})

    # Add the mosaic to a list only if the collection has images
    return ee.List(ee.Algorithms.If(filtered_day.size(), newlist.add(image), newlist))

daily_mosaics = ee.ImageCollection(ee.List(list_of_days.iterate(create_mosaics, ee.List([]))))

# Only keep days where area coverage is at least min_aoi_coverage
if aoi.size().getInfo() == 1:
    aoi_footprint = aoi.first().geometry()
else:
    aoi_footprint = ee.Feature(aoi.union(100).first())

def add_coverage(image):
    footprint = ee.Feature(image.get('footprint'))
    image_footprint = footprint.geometry()
    aoi_area = aoi_footprint.area()
    intersect = aoi_footprint.intersection(image_footprint, ee.ErrorMargin(1))
    intersect_area = intersect.area()
    coverage = ee.Number(intersect_area.divide(aoi_area))
    return image.set('AOI_COVERAGE', coverage)

daily_mosaics_aoi = daily_mosaics.map(add_coverage)
final_imagery = daily_mosaics_aoi.filter(ee.Filter.gte('AOI_COVERAGE', min_aoi_coverage))

# Apply cloud and water mask
def mask_s2_image(image):
    # Cast
    image = ee.Image(image)
    scl = image.select("SCL")  # Scene classification map
    # Mask out satured or defective (1), cloud shadows (3), water (6), clouds medium prop (8), clouds high prob (9)
    mask = scl.neq(9).And(scl.neq(8)).And(scl.neq(6)).And(scl.neq(3).And(scl.neq(1)))
    return image.updateMask(mask)


```

```python
masked_final_imagery = final_imagery.map(mask_s2_image)


# Extract vegetation indices (VIs) and relevant bands from the masked imagery, use separate functions for VIs that only use 10m imagery vs ones that use 20m
def add_vis_10m(image):
    # NDVI
    VIs = image.normalizedDifference(['B8', 'B4'])
    VIs = VIs.select("nd").rename("NDVI")
    # GCVI
    VIs = VIs.addBands(image.expression(
    '(NIR / GREEN) - 1', {
      'NIR': image.select('B8'),
      'GREEN': image.select('B3')
    }).rename('GCVI'))
    # EVI (note, bands are scaled by 10,000 which needs to be removed to get surface reflectance values)
    VIs = VIs.addBands(image.expression(
    '2.5*((NIR - RED) / ((NIR + 6 * RED - 7.5 * BLUE) + 1))', {
      'NIR': image.select('B8').divide(10000),
      'RED': image.select('B4').divide(10000),
      'BLUE': image.select('B2').divide(10000)
    }).rename('EVI'))
    # B4
    VIs = VIs.addBands(image.select('B4'))
    # B8
    VIs = VIs.addBands(image.select('B8'))
    return VIs.set({'Date': image.get("Date")})

def add_vis_20m(image):
    # Red-edge NDVI (Band 5 is 20m)
    VIs = image.normalizedDifference(['B8', 'B5'])
    VIs = VIs.select("nd").rename("reNDVI")
    # MTCI
    VIs = VIs.addBands(image.expression(
    '(NIR - RE) / (RE - RED)', {
      'NIR': image.select('B8'),
      'RE': image.select('B5'),
      'RED': image.select('B4')
    }).rename('MTCI'))
    #SeLI: Pasqualotto, N., Delegido, J., Van Wittenberghe, S., Rinaldi, M., & Moreno, J. (2019). Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). Sensors (Basel, Switzerland), 19(4), 904. doi:10.3390/s19040904
    VIs = VIs.addBands(image.normalizedDifference(['B8A', 'B5']).rename('SeLI'))
    #LAIgreen
    VIs = VIs.addBands(image.expression('5.405 * ((R865 - R705) / (R865 + R705)) - 0.114', {
        'R865': image.select('B8A'),
        'R705': image.select('B5')
        }).rename("LAIgreen"))
    # B5
    VIs = VIs.addBands(image.select('B5'))
    # B8A
    VIs = VIs.addBands(image.select('B8A'))
    return VIs.set({'Date': image.get("Date")})

vegetation_indices_10m = masked_final_imagery.map(add_vis_10m)
vegetation_indices_20m = masked_final_imagery.map(add_vis_20m)

# Calculate zonal stats for each VI/band
meters = 10
def zonalStats(image):
    date = image.get("Date")
    toReturn = image.reduceRegions(reducer=ee_reducer, collection=plot_boundaries, scale=meters)
    return toReturn.set('Date', date)

zs_10m = vegetation_indices_10m.map(zonalStats)
meters = 20
zs_20m = vegetation_indices_20m.map(zonalStats)
```

```python


#  Remove geometry from the zonal stats
def processFeature(feature):
    return feature.setGeometry(None)

def removeGeometry(featureCollection):
    fc = ee.FeatureCollection(featureCollection)  # Cast
    fc_date = ee.Date(fc.get('Date')).format('yyyy-MM-dd')  # Get date to assign
    fc_no_geometry = fc.map(processFeature)
    toReturn = fc_no_geometry.map(lambda x: x.set({'Date': fc_date}))
    return toReturn

zs_10m_no_geom = zs_10m.map(removeGeometry).flatten()
zs_20m_no_geom = zs_20m.map(removeGeometry).flatten()


#  Merge all of the data together
def cleanJoin(feature):
    return ee.Feature(feature.get('primary')).copyProperties(feature.get('secondary'))

filter = ee.Filter.equals(leftField = 'system:index', rightField = 'system:index')
simpleJoin = ee.Join.inner('primary', 'secondary')
zonal_stats = simpleJoin.apply(zs_10m_no_geom, zs_20m_no_geom, filter).map(cleanJoin)

# Export the data to Google Drive
task = ee.batch.Export.table.toDrive(collection=zonal_stats, description=output_file,
                        fileFormat='CSV', fileNamePrefix=output_file,
                        folder=output_folder)

task.start()
```