Workflows for Spatial Ex Ante Economic Analysis of Agronomy Interventions

Version 0.1

November 2024

INITIATIVE ON
Excellence in Agronomy

Maxwell Mkondiwa[a*], Jordan Chamberlin[b], Oyakhilomen Oyinbo[c]
a International Maize and Wheat Improvement Center (CIMMYT), Delhi, India
b International Maize and Wheat Improvement Center (CIMMYT), Nairobi, Kenya
c International Maize and Wheat Improvement Center (CIMMYT), Abuja, Nigeria
* Corresponding author: M. Mkondiwa (m.mkondiwa@cgiar.org)

INITIATIVE ON Excellence in Agronomy

Table of contents

INITIATIVE ON Excellence in Agronomy

Preface

This manual provides workflows for conducting spatial ex ante economic analysis of agronomic interventions. The statistical, mathematical and economic concepts in these workflows are already available in different books and journal articles. We especially leveraged the classic text by Alston et al (1995) that summarized all the methods for evaluating agricultural research and development innovations. Our task was simply to organize the different methods that spawned after Alston et al (1995) as standard operating procedures (SOPs) —i.e., analytical procedures that can be reproducible and replicated in other contexts. We also focused on how to transition from country level analyses that characterized that era to spatially gridded applications that recognize the heterogeneity associated with smallholder farming systems. In this manual, you will learn how to conduct spatially explicit (mostly gridded) economic evaluation of agronomic innovations with R and other programming software including Octave, and python.

Our purpose for these SOPs is to standardize the spatial ex ante economic analysis process with the aim of reducing the amount of time researchers spend on coding the common building blocks of spatial ex ante workflows and improving their efficiency.

The choice of which methods to showcase in the SOP reflects our primary goal which is to provide a simple and easy-to-use prioritization assessment framework for potential agronomic investments, which can be applied by economists, non-economists, project managers, and students.

We focus on three indicators for prioritization: yield gains, profit gains and robustness to risk. These have been recognized as some of the key indicators for assessing agronomic gains to technology.

1. Introduction

1.1. Overview

The workflows manual for spatial ex ante economic analysis consists of a suite of validated toolkits that use economic theory, econometrics, and mathematical optimization models to target investments of different agronomic innovations at disaggregated spatial scales, as well as evaluate the returns on these investments. The returns are expressed depending on the toolkit as yield gains, profit gains, probability of getting a threshold level of yield gains or profit gains, yield or profit risk, producer surplus, financial indicators like cost-benefit ratio and net present value, and willingness to pay measures.

We start with an exploratory toolkit which involves a literature review, back-of-the-envelope calculations, and stakeholder engagements. In the standard procedures for evaluating agricultural research (as described in Alston et al 1995)[1], this procedure involves using scoring and other short cut approaches. This procedure is not a replacement for the other more objective and data-grounded procedures we discuss next. It is a starting point for doing spatial ex ante work as it guides the nature and scope of toolkits to use next.

The next toolkit in the system assesses whether there are substantial economic gains for a farmer who is either risk agnostic or risk averse to likely adopt the technology. This toolkit assumes that the farmers are individually too small to affect the prices and quantities of other farmers.

In the next toolkit, we relax these assumptions to estimate the producer and consumer surplus while considering farmers' demand and supply price behavior and equilibrium relationships. The two toolkits presented assume that the technology already exists and that all that remains is to increase its adoption.

In the final toolkit, we present a case of first ascertaining whether there is adoption potential given spatially explicit endowments and whether by adding attributes to the technology, farmers are then more likely to adopt it. We then use that system to compute the economic value of that adoption to the farmers.

How does one choose which toolkits to use? We present next a decision tree for choosing the analytical methods given the available data, expertise and time.

1.2. Decision tree to spatial ex ante economic analysis toolkits

This manual has provided a list of spatial quantitative ex ante toolkits that are used to prioritize potential investments. These toolkits mostly use existing survey data or baseline data that most projects ideally collect prior to implementation.

[1] Within CSISA and EiA, the CAPTAIN tool (now called PAiCE) is a clear example of this toolkit.

Figure 1 shows a decision tree for selecting which ex ante approach to use for a selected study. We categorize the decision steps into four layers. First, one has to conduct a literature review, back-of-the-envelope calculations of structural changes in the economy, and stakeholder engagements. This layer needs to be done regardless of the comprehensive approaches that are later used in the spatial ex ante assessment. It is possible to stop and start implementing at this point if prior ex ante studies were already conducted on the topic of interest and at a sufficient scale. In the second layer, one gathers all the necessary datasets required for the ex-ante work. If there is no data, then instead of scaling the interventions, it is best to work with stakeholders to design on-station, on-farm experiments, quantitative and qualitative surveys to start gathering evidence to be used for ex ante analysis.

If survey data exist already, then one needs to start with the spatial profitability and risk assessment toolkit. In this toolkit, the researcher needs to ask if the technology in question is sufficiently studied elsewhere such that there are already functional forms to use the parametric approach. If not clear on this then, he/she may use the causal ML based approaches. The researcher may use the spatial Bayesian kriging approach if the targeting is to focus on locations to implement including out of sample. He/she may consider the policy learning optimization approach if he/she is interested in understanding the indicators to use when partitioning who needs to be prioritized beyond the spatial aspects.

If crop simulation and long-term experimental datasets are available, one can use the spatial profitability and risk assessment toolkit as well. But in addition, he/she may be interested in using the evidence from these experiments to suggest new combinations of treatments that should be tested or scaled beyond what is observed to be beneficial. For that, the researcher can use the modern portfolio approaches (i.e., mean-variance optimization or mean-conditional value at risk optimization). If instead he/she is just interested in recommending for each grid the most robust practice for scaling, then he/she can use the willingness to pay bounds approach which recommends the best practice for any risk-averse farmer to find it optimal to follow that strategy.

The spatial profitability assessment toolkit and the spatial optimization toolkit will provide an individual assessment of the benefits. However, one may be interested to understand if this makes sense socially as well given the demand and supply behavioral patterns (i.e., elasticities). The equilibrium displacement modelling framework, also called economic surplus approach (Alston et al 1995) is the traditional way of assessing producer and consumer surplus for the new agronomic technology. This approach utilizes supply/demand elasticities, the spatially explicit yield gains and cost reduction estimates and the assumed adoption trends to evaluate the returns on investments at a disaggregated level.

What if the new technology is just a variant of old technology. For example, a new mix of herbicides or a new variety. In most cases, these are not widely adopted to warrant a spatial profitability assessment using observational data. In addition, the performance in agronomic trials will likely be misleading because farmers have not yet learned how to use the technology appropriately. How do we predict whether this new technology will be adopted and that it’s worth investing in? For this, the researcher needs to consider the structurally differentiated agronomy toolkit which uses characteristics of the technology and location-specific characteristics (including farmer demographics) to predict whether the new technology embodying particular traits is worth investing in.

As can be seen from the decision tree, these toolkits can be used in tandem because they give different insights into the likely benefits of investing in a particular technology. In addition, these approaches are not exhaustive. Each of the workflows we have presented has several variants and complementary methods that researchers can explore.

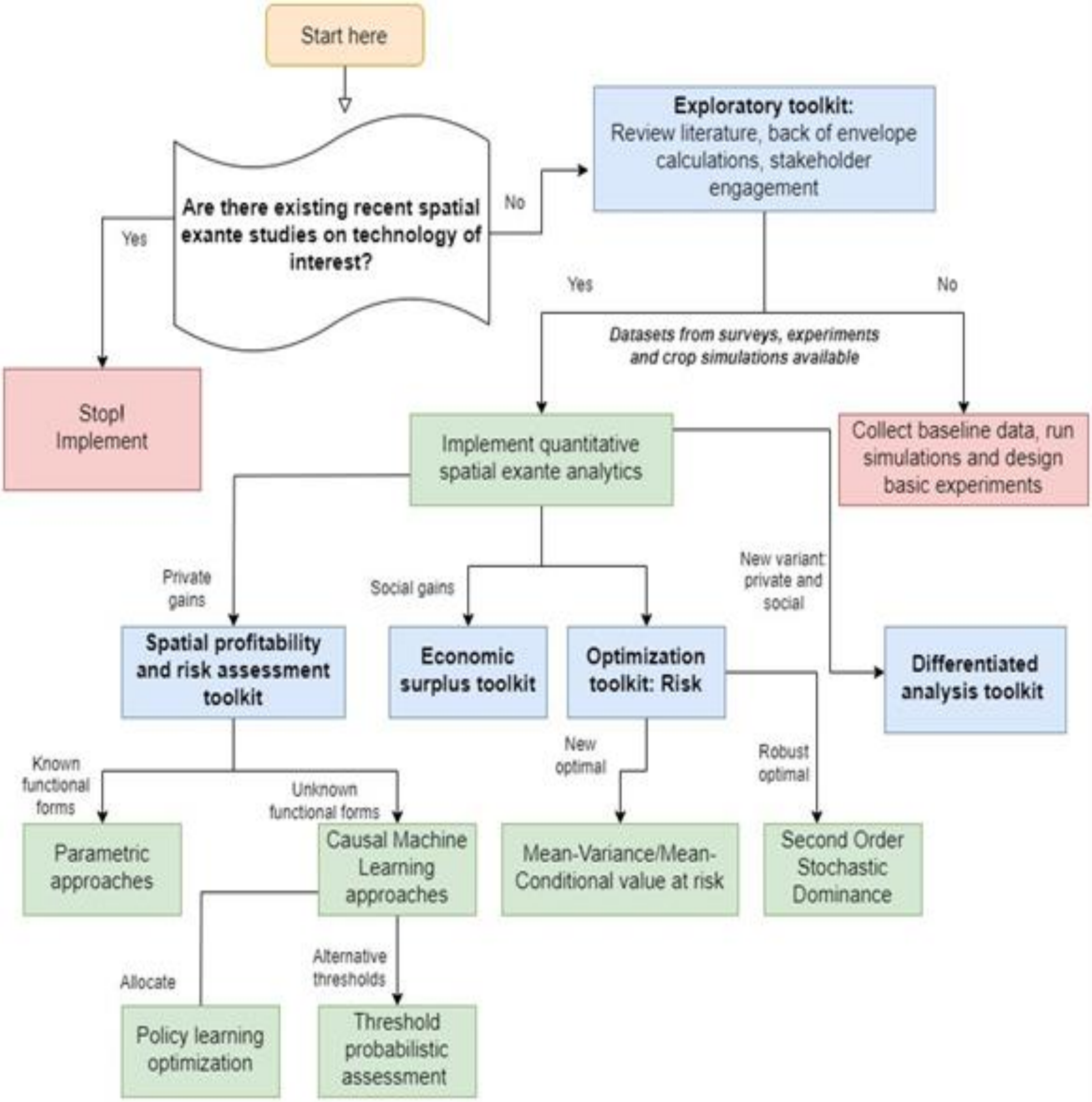

Figure 1: Decision tree for selecting spatial ex ante approaches

1.3. Target audience

The toolkits can be used by researchers, project managers, and students to conduct spatial ex ante analyses that guide investments. Our target audience are researchers who are tasked to conduct spatially explicit ex ante analyses. We thus assume that the researchers have used most of these tools before or are at least aware of them. For those not familiar with the methods, we encourage them reader to go through the suggested references.

1.4. Stylized example: Sowing date use case

We focus on sowing date use case being implemented in India as an exemplar for the ex ante analytics presented in this SOP. However, we have applied the same techniques for other use cases including: (1) Herbicide integrated weed management in wheat, (2) multiple irrigations in wheat (irrigation scheduling) and short- and long-duration wheat varieties (varietal choice).

1.5. Replication materials

The replication materials for all the toolkits can be accessed on EiA Ex-ante github page: https://github.com/EiA2030-ex-ante. We use publicly available datasets and one should be able to replicate all the results in this SOP.

2. Exploratory toolkit

Purpose

There are cases in which spatial ex ante economic analysis can be done using theory-grounded back-of-the-envelope calculations (e.g., Binswanger 1986), literature review (e.g., Pardey et al 2016) or stakeholder prioritization workshops (e.g, Lee et al 2014). We call these “exploratory” workflows as they are commonly used to guide more comprehensive methods. The cases in which these exploratory methods are most appropriate include when there is ample evidence on returns on investments (ROI) for that particular technology, or when the time and resources required would not allow collection of the necessary data for the comprehensive analytics. We discuss in this section the merits and demerits of the approach, the data requirements and links to replication materials.

Advantages

- It is the simplest approach and has less data requirements.
- It is useful when there is a lack of data for a formal quantitative evaluation.
- For some agricultural innovations, formal quantitative methods are difficult or even impossible to implement.
- When there is mixed evidence on the probability of success, a stakeholder engagement can help in understanding the likely values. In addition, it helps in building an institutionalized “economic way of thinking” (Alston et al 1995).
- Less time consuming.

Disadvantages

- The approach is less precise and can be biased by the nature of evidence available and by the set of stakeholders engaged.

Stylized use case: Where to target sowing date advisories?

Using literature review, it is clear that early sowing of wheat has yield advantages. In focus group discussions, farmers also expressed the same.

Input data requirements

- Literature database with estimates for each location
- Stakeholder workshop with a scoring matrix

Workflow

Following Alston et al (1995) and Lee et al (2014), the scoring toolkit involves the steps shown in figure 2. As can be seen from figure 2, the approach involves triangulating across three different approaches: scoring (Alston et al 1995, Lee et al 2014), literature reviews (e.g., Pardey et al 2016), and theory-driven back-of-the-envelope calculations (Binswanger 1986, Goldman 1993).

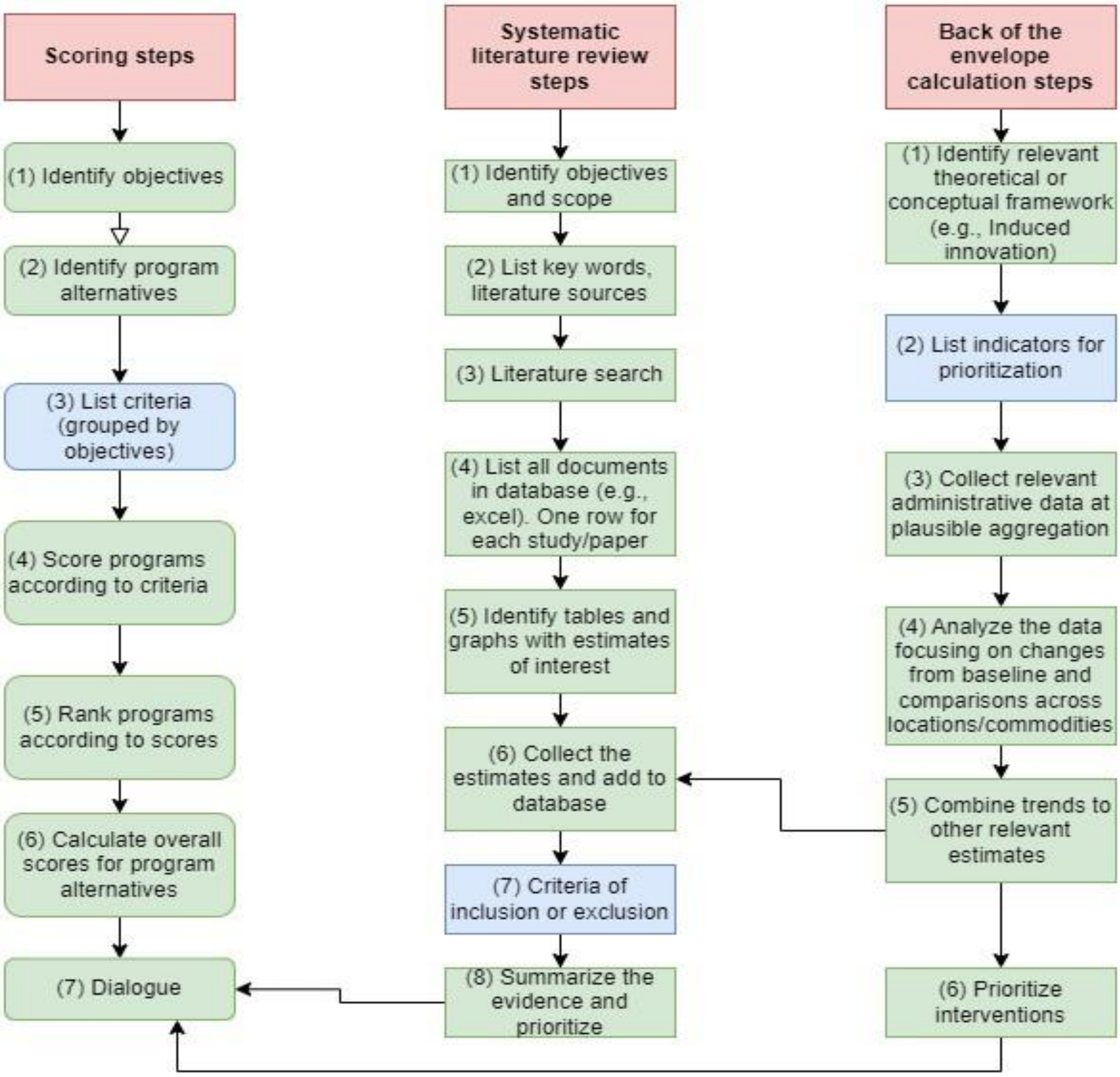

Figure 2: Stylized workflow for literature review, scoring and back of the envelope calculations

Box 1 shows the mathematical basics underpinning scoring models, systematic literature reviews, and back-of-envelope calculations. The goal of these approaches is to approximate net efficiency gains to agronomic R&D investments by approximating sufficient statistics for approximating returns to investments. These include expected aggregate area that will adopt the new agronomic intervention, expected proportional yield gains, and expected value of production given prevailing prices and production. The approximations can be made spatially explicit if stakeholders are able to provide expert opinions at a disaggregated level.

Box 1: Mathematical basis for scoring models

Following Alston et al (1995, p.467), gross (G) and net (N) efficiency gains are key outputs of scoring exercises. These can be expressed in ordinal or cardinal form depending on the availability of the data. This is defined as:

$$G_i = A_i^{MAX} p_i E\left(Y_i^{MAX}\right) P_i Q_i$$
$$N = \frac{G}{R}$$

where

P: Baseline price
Q: Baseline quantity
E(Y): Proportional yield increase due to the technology
P: Probability of success
A: Proportion of farmers likely to adopt
R: Research costs

Replication materials

https://github.com/EiA2030-ex-ante/Ex-ante-Summary-Tool

Key references

Alston, J.M., Norton, G.W., and Pardey, P. 1995. "Science under scarcity: Principles and practice for agricultural research evaluation and priority setting". CAB International. See Chapter 7 (pp. 463-498).

Binswanger, H.P. 1986. "Evaluating Research System Performance and Targeting Research in Land-abundant Areas of Sub-Saharan Africa." *World Development* 14(4): 469-475. Doi: https://doi.org/10.1016/0305-750X(86)90063-X.

Goldman, A. 1993. "Agricultural Innovation in Three Areas of Kenya: Neo-Boserupian Theories and Regional Characterization." *Economic Geography* 69(1): 44-71. Url: http://www.jstor.org/stable/143889.

Lee, D.R., Edmeades, S., Nys, E., McDonald, A., Janssen, W. 2014. "Developing local adaptation strategies for climate change in agriculture: A priority-setting approach with application to Latin America". *Global Environmental Change* 29: 78-91. Doi: https://doi.org/10.1016/j.gloenvcha.2014.08.002.

Pardey, P.G., Andrade, R.S., Hurley, T.M., Rao, X., and Liebenberg, F.G. 2016. "Returns to food and agricultural R&D investments in Sub-Saharan Africa, 1975-2014". *Food Policy* 65: 1-8. Doi: https://doi.org/10.1016/j.foodpol.2016.09.009.

3. Spatial ex ante profitability and risk toolkit

3.1. Spatial parametric and non-parametric production function approach with risk

Purpose

The conventional approach to spatial targeting is to estimate a parametric production function (also called crop response function), then use profit maximization or its dual cost minimization to identify optimal demand for the associated technology. The parametric production function can be linear (e.g., quadratic production function) or non-linear (e.g., quadratic and plateau function). These approaches have been extensively used in the literature as such we do not cover them extensively here. The conventional approach has recently been extended to non-parametric settings (see Palmas and Chamberlin 2020 using random forest estimators). If risk is considered important, the traditional approach is to assume a quadratic utility function and use mean-variance approaches to assess optimal choices under risk aversion preferences (e.g., Just-Pope production function or moments productions). To use this approach for spatial ex ante, one can simply use spatial Bayesian models for point-referenced data (Note: spatial econometric approaches can also be used for this extension to the traditional model).

Advantages

- Simple to use with standard econometric approaches (e.g., ordinary least squares estimator).

Disadvantages

- Difficult to identify the appropriate functional form and the results are largely dependent on this choice.
- Difficult to project into scenarios that have not been observed yet.

Stylized use case: Are sowing date advisories risk proof?

We use CSISA-KVK trial data to understand whether early sowing of wheat and planting of long duration wheat varieties increase mean yield and reduce risk. Workflow Figure 4 shows a workflow for the spatial parametric production function approach. This is categorized into four steps. First, one estimates a production risk function model using either the residual-based (e.g., Just-Pope production function) or the moments-based approach using ordinary least squares approach. If there are concerns with endogeneity, then one can correct for these using the instrumental variables approach or other quasi-experimental methods. The simple approach we recommend is the approach proposed in Lewbel (2012). Third, to make the estimate spatially explicit, we recommend using a spatially varying coefficient model to get estimates for each pixel in the area of interest. Finally, one can use input and output prices, especially spatially varying prices to create economic indicators of interest.

Workflow

Figure 4 shows a step-by-step workflow for implementing the parametric production function and risk approach. The approach involves estimating a set of Just-Pope production (mean) and risk models (variance) or moments production (mean) and risk (variance, skewness, kurtosis) models.

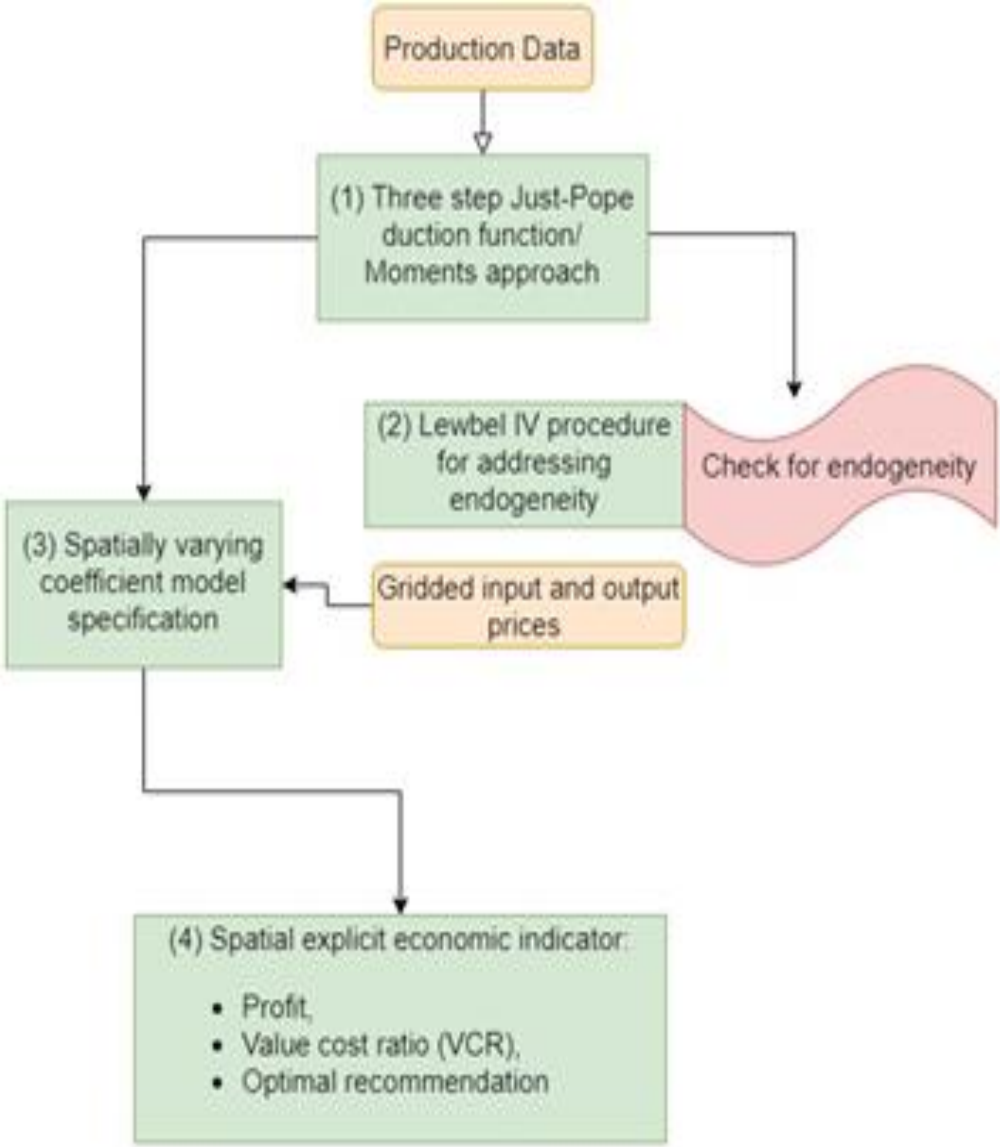

Figure 4: Workflow for spatial risk production function

Armed with these estimates which can be spatially explicit if a spatially differented regression approach is used, one then estimates key indicators of economic welfare including profits, value cost ratios, risk premiums and certainty equivalents. Box 2 shows how to derive the risk premiums and certainty equivalents.

Box 2: Mathematical basis for parametric production function and risk models

For a risk model based on the moments approach, we define a stochastic production function as follows (see Di Falco and Chavas 2009):

$$y = g(x, v) = f_1(x, \beta_1) + \mu$$

Where $f_1(x, \beta_1) \equiv E[g(x, v)]$ is the mean of $g(x, v)$ and $\mu \equiv g(x, v) - f_1(x, \beta_1)$ is a random variable with mean 0.

For the traditional production function (with profit maximization as the objective), the optimal solution is found at $\frac{\partial y}{\partial x} = \frac{r}{p}$ where p is the output price while r is the input price.

For risk models, Di Falco and Chavas (2009) further define higher moments of $g(x, v)$ as

$$E\{[g(x, v) - f_1(x, \beta_1)]^k\} = f_k(x, \beta_k)$$

For $k = 2,3,4$ ... where $f_1(x, \beta_1)$ is for the first central moment, $f_2(x, \beta_2)$ is for the second central moment (variance), $f_3(x, \beta_3)$ for the third central moment (skewness) and $f_4(x, \beta_4)$ for the fourth central moment (kurtosis).

We follow Di Falco and Chavas (2009) and Dong and Mitchell (2023) in evaluating the cost of risk using risk premium calculated as follows:

$$R = R_2 + R_3 = \frac{1}{2} r_2 \pi_2 + \frac{1}{6} r3 \pi_3$$

Where $r_2 = -\left(\frac{\frac{\partial^2 U}{\partial \pi^2}}{\frac{\partial U}{\partial \pi}}\right)$, and $r_3 = -\left(\frac{\frac{\partial^2 U}{\partial \pi^2}}{\frac{\partial U}{\partial \pi}}\right)$.

Welfare change is defined by the certainty equivalent

$$CE = \pi_1 - R$$

Using the constant relative risk aversion (CRRA) utility function:

$$U(\pi) = \begin{cases} \dfrac{\pi^{1-\theta}}{1-\theta} & \text{for } \theta \geq 0 \text{ and } \theta \neq 1 \\ \ln \pi & \text{for } \theta = 1 \end{cases}$$

$r_2 = \frac{\theta}{\pi_1}$ and $r_3 = -\frac{\theta(\theta+1)}{\pi_1^2}$

where: $R = R_2 + R_3 = \frac{1}{2}\frac{\theta}{\pi_1} \pi_2 + \frac{-\theta(\theta+1)}{6\pi_1^2} \pi_3$

For the cases where kurtosis is also computed, we can follow Shi et al (2013) and Mourtzinis et al (2023) to include additional terms in the equation above as follows:

$$R = R_2 + R_3 + R_4 = \frac{1}{2}\frac{\theta}{\pi_1} \pi_2 + \frac{-\theta(\theta+1)}{6\pi_1^2} \pi_3 + \frac{\theta(\theta+1)(\theta+2)}{24\pi_1^2} \pi_4$$

According to Li et al (2021) and Gollier 2004, θ is usually between 1 and 5.

Stylized outputs

Figure 5 shows the Just-pope production and risk function results while figure 6 shows the moments production and risk function results. In both cases, late sowing of wheat (e.g., T5 corresponding to after December 15) has higher yield losses and higher downside risk.

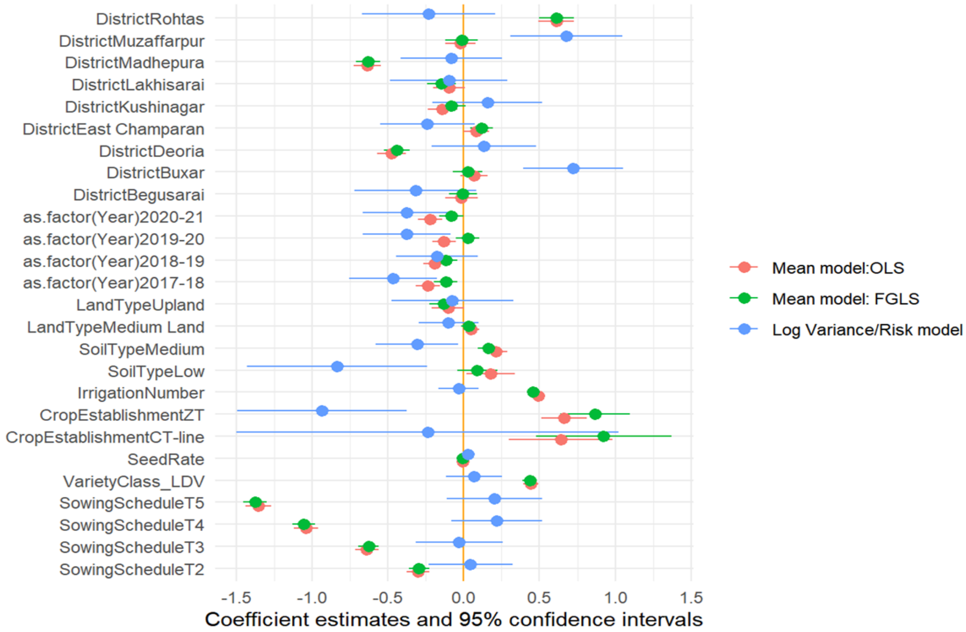

Figure 5: Just-Pope production risk model

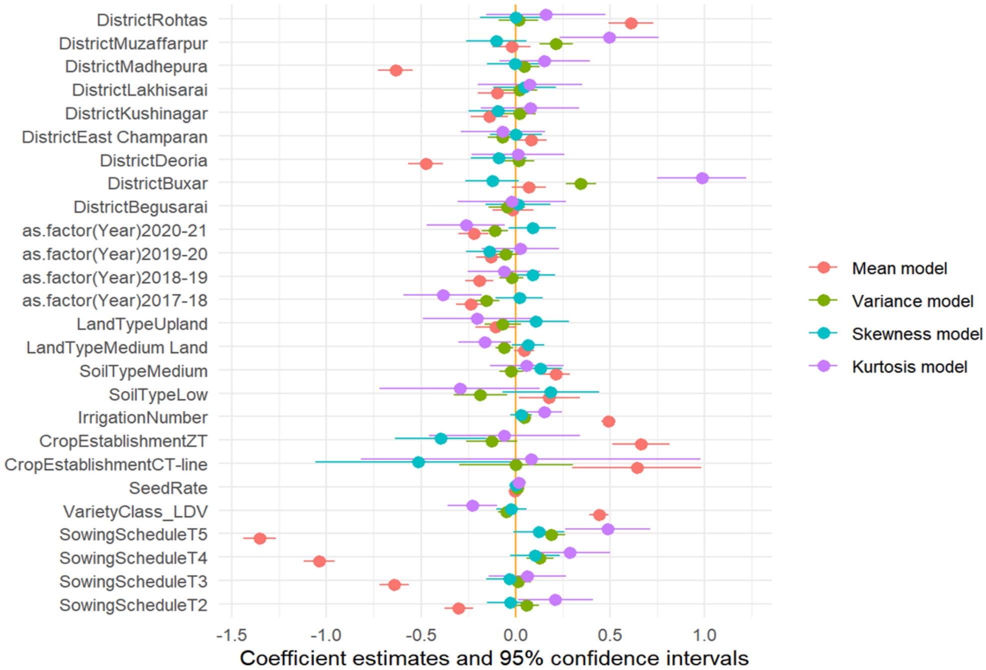

Figure 6: Moments based production risk model

Replication materials

https://eia2030-ex-ante.github.io/SpatialParametricProduction_Risk_Model/ and https://github.com/EiA2030-ex-ante/Nigeria_response_uncertainty.

Key references

Antle, J.M. 1983. “Testing the stochastic structure of production: A flexible moment-based approach”. *Journal of Business and Economic Statistics* 1(3): 192-201. Doi: 10.1080/07350015.1983.10509339.

Di Falco, S., and Chavas, J. 2009. “On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia”. *American Journal of Agricultural Economics* 91(3): 599-611. Doi: https://doi.org/10.1111/j.1467-8276.2009.01265.x.

Lewel, A. 2012. Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics* 30: 67–80. Doi: https://doi.org/10.1080/07350015.2012.643126

Mourtzinis, S., Mitchell, P., Esker, P., Cerrudo, A., Naeve, S., and Conley, S. 2023. “Field-level yield benefits and risk effects of intensive soybean management across the US”. *Field Crops Research* 301: 109012. Doi: https://doi.org/10.1016/j.fcr.2023.109012.

Palmas, S., and Chamberlin, J. 2020. “Fertilizer profitability for smallholder maize farmers in Tanzania: A spatially-explicit exante analysis”. PLoS ONE 15(9): e0239149. Doi: https://doi.org/10.1371/journal.pone.0239149.’

Shi, G., Chavas, J., and Lauer, J. 2013. “Commercialized transgenic traits, maize productivity and yield risk”. *Nature Biotechnology* 31: 111-114. Doi: https://doi.org/10.1038/nbt.2496.

3.2. Causal ML and spatial probabilistic assessment model

Purpose

In some cases, the farmer is not only interested in shifting to a technology that gives the highest yield gains, but also the one that has the highest chance of giving him/her yields beyond a particular threshold.

Advantages

The spatial probabilistic approach adds value under the following circumstances:

- One is interested in segmentation of zones of opportunities.
- One is interested in threshold probabilities as measures of uncertainty.

Disadvantages

- The spatial Bayesian models are computationally expensive, especially for large N data and can take many weeks to produce results. This can be resolved by using High Performance Computers.

Stylized use case: Where to target sowing date advisories that achieve yield gains beyond a particular threshold?

A farmer requires a substantial yield gain to change from the conventional behavior. In recommending planting date changes, it is therefore important to provide the confidence we have that the farmer will likely attain yields higher than that threshold. A probabilistic assessment approach allows this through a threshold probability—the probability that a farmer in that location will achieve yield gains above the threshold.

Input data requirements: The approach requires geo-referenced farm plots with attendant yield and traditional production variables (e.g., seed, fertilizer, weed management, and other agronomic management variables).

Toolkit workflow

Machine learning and spatial models have been criticized for focusing on prediction, leaving aside questions of causality. For a spatial ex ante analysis, there is also a question of the predictions incorporating optimizing behavior as such one needs to use causal estimates to make predictions. We use the latest methods in causal machine learning like causal random forest (Athey et al. 2019) to analyze the impact of irrigation schedules on yield. These models have also been applied by McCullough et al (2022) for fertilizer crop response modelling. The key limitation of causal random forest algorithm is that it neglects dependent processes (Saha et al 2021). To consider the spatial correlation, we use a Bayesian spatially varying coefficient model in the second stage. One can also use a hybrid ML and spatial model called random forest GLS (Saha et al 2021). This toolkit is implemented by following the steps shown in figure 7.

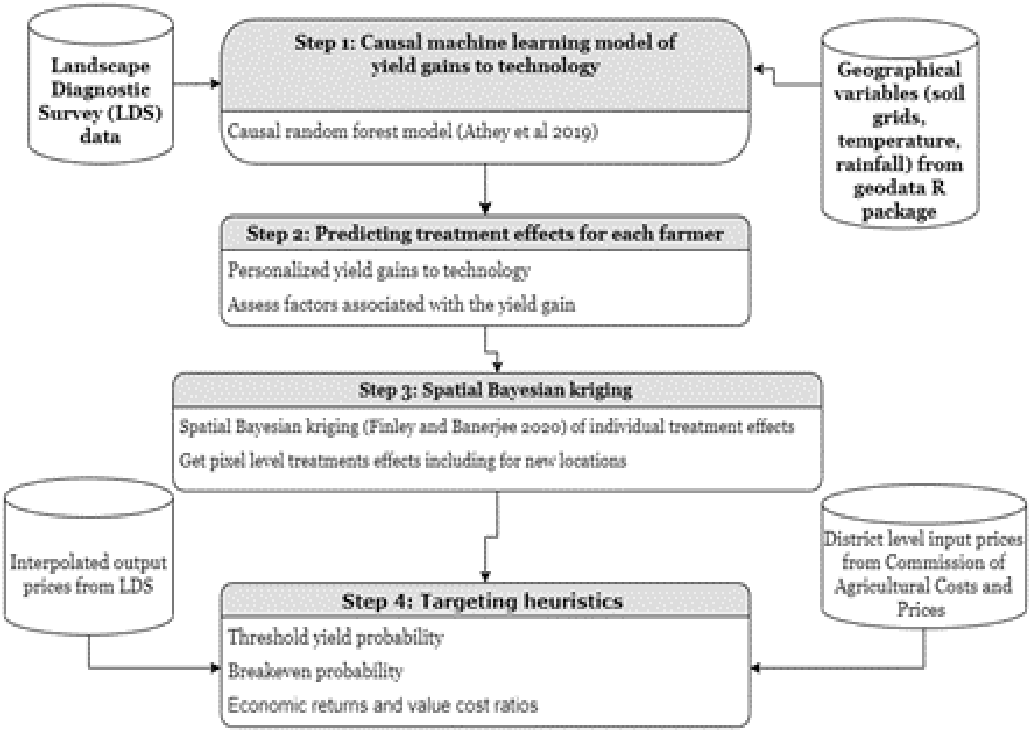

Figure 7: spatial probabilistic assessment toolbox

The mathematical details of each of the modeling steps in the figure are presented in Box 2.

Box 3: Mathematical basis for causal ML prioritization models

Step 1: [Causal random forest model]—*Estimate a yield crop response model using causal random forest model with augmented inverse weighted propensity score (AIWP). Causal random forest model can be written as extension of heterogeneity analysis using interactions,*

$$y = \mu(x) + \tau(X) \times W + \epsilon$$

Where y is wheat yield (tons/ha), $u(x)$ is the control effect (that is assuming farmer is applying only one irrigation), W is the treatment dummy ($W = 1$ if treated meaning applied more than 1 irrigation, 0 otherwise), $\tau(x)$ is the conditional treatment effect, and ϵ is the error term. Following the potential outcomes framework, the conditional average treatment effect can be expressed as

$$\hat{\tau}_i(x) \equiv E(Y_i(1) - Y_i(0)|X_i = x)$$

Unlike random forest models, causal random forests use honest splitting in which splitting is done on training sample and uses independent sample to estimate leaf means. Causal random forests also allow orthogonalization for observational (survey) data using Inverse-Probability Weights predicted *from probit model. Augmented Inverse-Probability Weighted Conditional Average Treatment Effect (AIPW-CATE) is given by*

$$\hat{\Gamma}_i := \hat{\tau}(X_i) + \frac{W_i}{e(X_i)}\left(Y_i - \hat{\mu}(X_i, 1)\right) - \frac{1 - W_i}{1 - e(X_i)}\left(Y_i - \hat{\mu}(X_i, 0)\right)$$

where $e(X_i)$ are treatment assignment probabilities. In all the specifications, we control for all confounders including variety type, date of sowing, fertilizer applied, soil quality variables including soil nitrogen and pH, precipitation, temperature, and farm size.

Step 2: Predict conditional average treatment effects (CATEs) for each location.

The advantage of causal random forest is that we can predict the conditional average treatment effects for each individual sampled farmer. This allows analyzing farmer heterogeneity across space and socio-economic groups.

Step 3: Spatial Bayesian kriging of heterogenous treatment effects.

The CATEs produced in step 2 are for the sampled farmers. What about those we haven't visited yet? For this, we used spatial Bayesian kriging to produce gridded maps showing yield responses, economic responses, and probabilities of getting gains above some threshold. This approach builds on an idea from the double machine learning approach in which the best linear projection of the conditional average treatment effect is used to analyze the effect of covariates on the conditional average treatment effect.

$$\tilde{\tau} = \beta_0 + \beta Z + \epsilon$$

where Z are covariates that explain the robust scores. We specifically extend this estimation by including a spatial effect as second stage Gaussian process (GP) using Spatial Bayesian Linear Model (Finley and Banerjee 2020):

$$\tau^{Score} = \beta_0 + \omega(s) + \epsilon,$$
$$\omega(s) \sim GP(0, C(.,.)$$
$$\epsilon \sim N(0, \tau^2)$$

where spatial effect ($\omega(s)$) which we model as an exponential Gausian Process (GP), i.e., $C(s_i, s_j) = \sigma^2\, exp(-\phi \left\|s_i - s_j\right\|_2)$. The exponential Gaussian Process is a function of the interlocation distance ($\left\|s_i - s_j\right\|_2$), the spatial variance (σ^2) also called partial sill, and the decay parameter (ϕ). The error process in equation is modelled as a normal distribution with mean 0 and non-spatial variance (τ^2) also called nugget.

The spatial Bayesian approach allows predicting treatment effects to new locations with parameter uncertainty. The posterior estimates are then used to calculate threshold probability—that is, the probability that yield gains will exceed a target yield change level, $\bar{y}$ (this may be a break-even yield). The threshold probability is given as:

$$Threshold\ probability = Pr[(\Delta y) > \bar{y}]$$

Step 4: Determine optimal (most beneficial) technology choice by comparing gridded yield returns (CATEs) and economic returns across technologies*. These are the spatially explicit yield and economic* *returns to agronomic innovations. Using the gridded yield gains (Δy), wheat prices (p^y), change in agronomic innovation costs ($\Delta f \times p^f$), we calculate economic returns (ER) and value cost ratio (VCR) as*

$$ER = \Delta y \times p^y - \Delta f \times p^f$$

$$VCR = \frac{(\Delta y \times p^y)}{\Delta f \times p^f}$$

These estimates can be compared or combined with agronomic trial estimates using Bayesian approaches as in Mkondiwa et al (2024).

See Athey et al (2019) for more details.

Stylized outputs

Using this toolkit, we see in Figure 8 that farmers in much of the area of interest (Bihar) would find early planting of wheat most beneficial and have a high probability of getting an additional 100kg/ha due to early sowing alone.

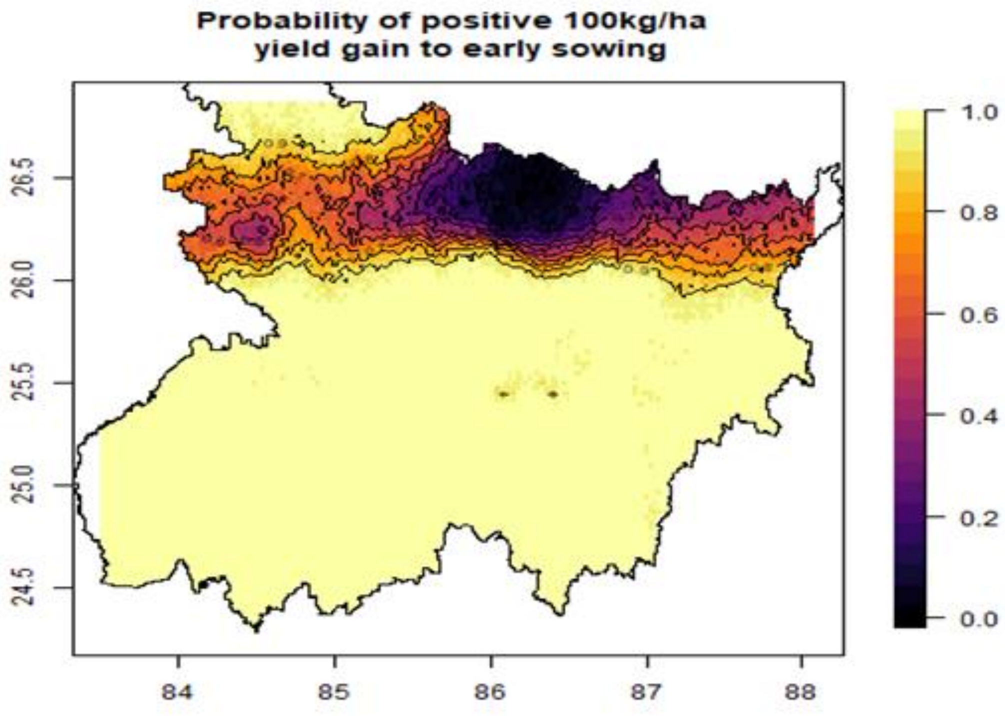

Figure 8: Stylized output for spatial probabilistic assessment showing the probability of yield gains of above 100kg/ha with early planting of wheat (i.e., before 21st Nov)

Replication materials

https://github.com/EiA2030-ex-ante/CausalML_SpatialBayes_Model. Other interpolations methods can be found here:https://github.com/EiA2030-ex-ante/Interpolation_Kriging_Methods_for_Exante

Key references

Athey, S., Tibshirani, J., and Wager, S. 2019. "Generalized Random Forests." *The Annals of Statistics* 47(2): 1148-1178. Doi: 10.1214/18-AOS1709.

McCullough, E.B., Quinn, J.D., Simons, A.M. 2020. "Profitability of climate-smart soil fertility investment varies widely across sub-Saharan Africa." *Nature Food* 3:275-285. Doi: https://doi.org/10.1038/s43016-022-00493-z.

Mkondiwa, M., Hurley, T.M., and Pardey, P.G. 2024. "Closing the gaps in experimental and observational crop response estimates: a bayesian approach". Q Open 4. Doi: https://doi.org/10.1093/qopen/qoae017.

3.3. Causal ML and policy learning optimization model

Purpose

To make individualized or personalized recommendations from observational data in a data-driven manner using causal machine learning frameworks.

Advantages

- Data-driven approach of recommending alternatives without making functional form assumptions. This is especially useful for agricultural inputs for which we do not have a clear functional form e.g., irrigation, sowing dates.

Disadvantages

- It requires enough sample sizes for each of the options being compared. This means that for new innovations which have not been extensively adopted, this approach would not be beneficial.

Stylized use case: Targeting sowing date advisories to individual farmers

While sowing date and many other recommendations are made based on climatic, biophysical and economic aspects, there may be several individual level reasons for not following the recommendations, e.g., labor shortage, as family members may be busy with other duties during those weeks. We propose a robust methodology that rests on causal machine learning and policy learning to make recommendations that are the most beneficial for each individual farmer.

Input data requirements: The data required is the same as for any conventional production function or impact assessment. These include yield, agronomic management variables (e.g., fertilizer applied), socio-economic variables, and input and output prices. One however, needs enough sample sizes for the treatment and control groups therefore the method works only for a technology which has been widely adopted.

Toolkit workflow

Figure 9 shows a step-by-step workflow for implementing the policy learning optimization model. The initial two steps are similar to the causal ML and probabilistic framework presented before. The main addition in this framework is to use policy tree algorithm to identify up to 3 variables (or farmer characteristics) than can be used when targeting the agronomic interventions. This can be desirable in that to enlist farmers to interventions which could be optimal for them, one does not need to collect all variables for the new sample. If the model is well calibrated, one needs only to ask those 3 questions to ascertain in a data driven manner if the interventions would be beneficial those particular farmers. Box 4 shows the mathematical basis for policy learning optimization part of the workflow.

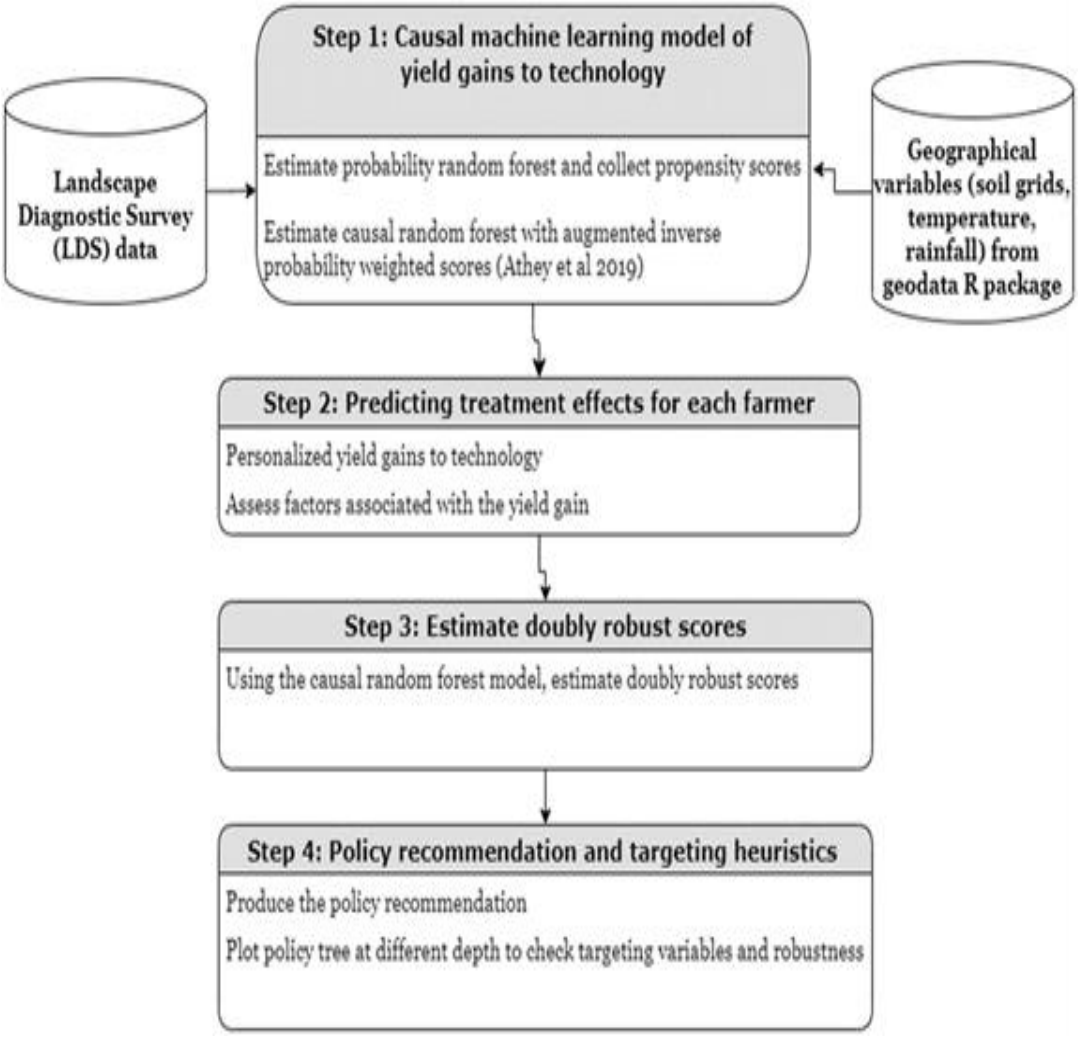

Figure 9: Workflow for causal ML and policy learning optimization

Box 4: Mathematical basis for causal ML and policy learning models

Given the causal ML results from Box 3, the goal is to choose a policy (a mapping from few observable characteristics to optimal practice) that maximizes the value function

$$\hat{\pi} = argmax \left\{ \hat{V}(\pi) : \pi \in \Pi \right\}$$

where

$$\hat{V}(\pi) = \frac{1}{n} \sum_{i=1}^{n} (2\pi(x_i) - 1)\, \hat{\Gamma}_i$$

$$\hat{\Gamma}_i = \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + \frac{W_i}{\hat{e}(X_i)}\big(Y_i - \mu_1(X_i)\big) - \frac{1 - W_i}{1 - \hat{e}(X_i)}\big(Y_i - \hat{\mu}_o(X_i)\big)$$

See Athey and wager (2021) for more details.

Stylized outputs

The outputs of steps 1 to 3 in the workflow are the individual level estimates of the yield gains from the proposed agronomic innovation as before. Figure 10 shows the distribution of yield gains to early sowing. Everyone in the sample would get a positive yield gain if they advance their planting strategy as compared to sowing after 16 December. The highest yield gains are with planting before 10^{th} November. However, the results in this figure do not prescribe a recommendation for that farmer. To prescribe a recommendation, we need to assume some objective function of the farmer. Policy learning uses minimum regret as an objective function to prescribe a best practice for each farmer. Figure 11 then shows the transition matrix from status quo to proposed agronomic practice for each individual farmer.

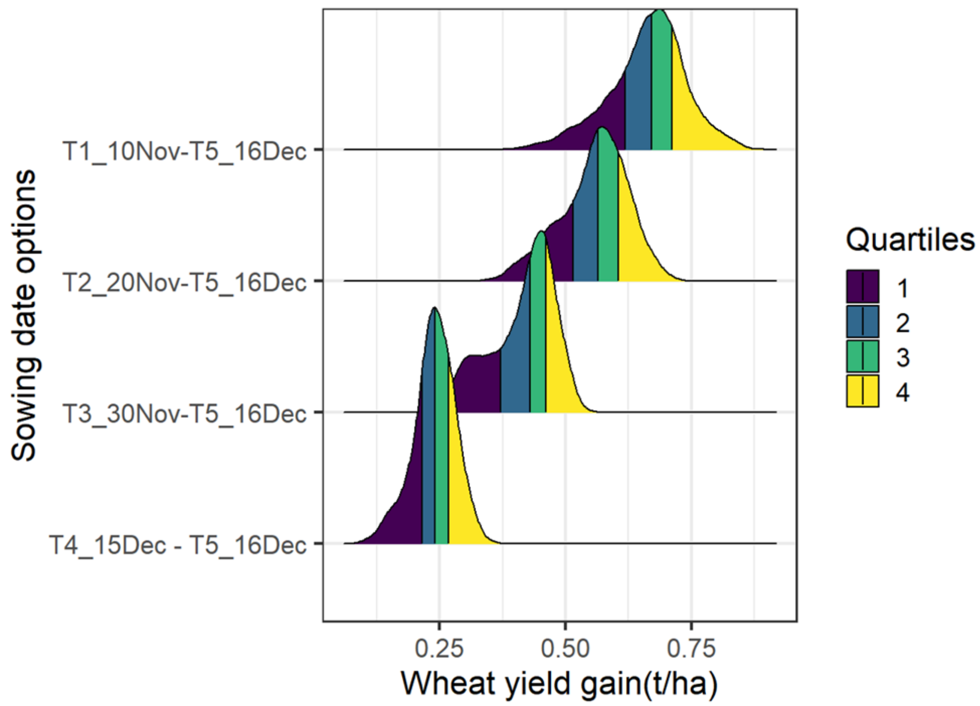

Figure 10: Distribution of conditional average treatment effects of wheat yield gains to early sowing from multi-armed causal ML model

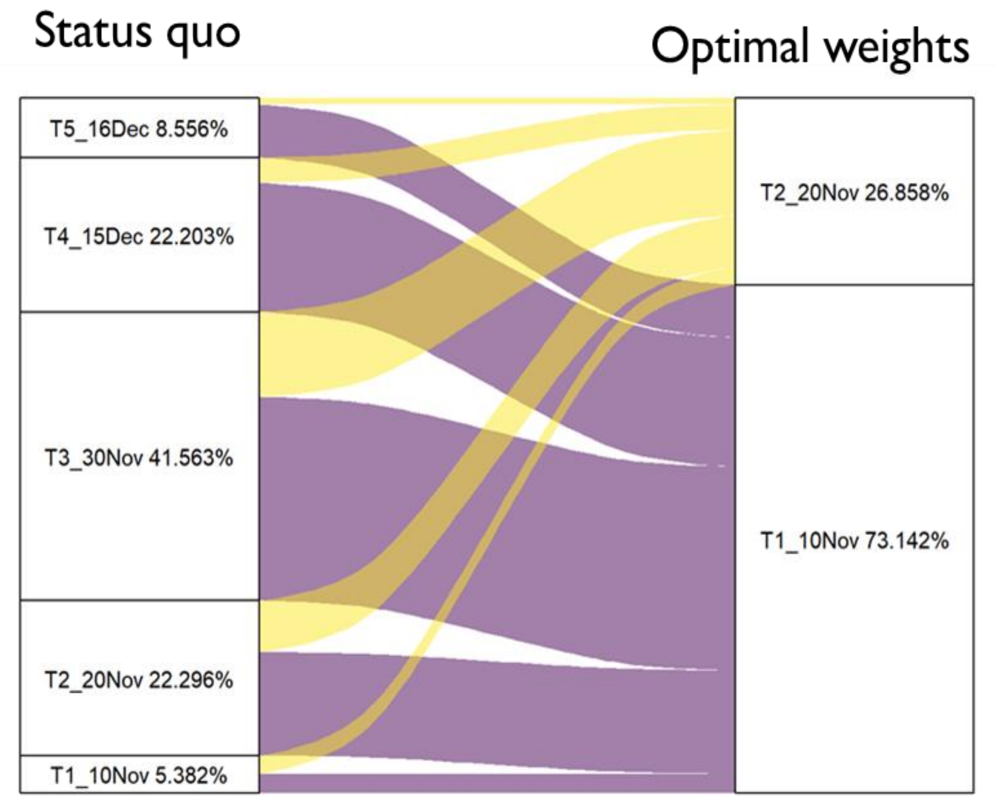

Figure 11: Transition matrix from status quo (as of 2019) to optimal allocations

Replication materials

https://eia2030-ex-ante.github.io/causal_RF_targeting/ for an application in India and https://github.com/MaxwellMkondiwa/TZ_Spatial_NUE_Trends_Causal_AI for an application in Tanzania.

Key references

Athey, S., and Wager, S. 2021. "Policy learning with observational data". Econometrica 89 (1): 133-161. Doi: https://doi.org/10.3982/ECTA15732.

4. Spatial optimization toolkit: Computational risk-return modeling

4.1. Mean-Variance (EV) and Mean-Conditional Value at Risk (CVaR) modern portfolio theory optimization

Purpose

Mean-variance (E-V) analysis seeks to maximize returns at the minimum risk (or variance) while it's variant, conditional value at risk (CVaR) also evaluates the downside risks. These two risk optimization models are part of the epoch of modern portfolio theory approach that was introduced by Harry Markowitz to identify efficient diversification options for investments. The approach can be novel in agronomic applications in that new combinations of treatments (aka portfolios) can be discovered for future testing and demonstrations (see Nayak et al 2024 for an example using long term experiment data of diversified cropping systems under conservation agriculture).

Advantages

- Allows selection of multiple alternatives beyond combinations observed in the data.

Disadvantages

- Focuses only two moments (mean and variance for EV analysis and mean and expected shortfall for CVaR) yet other moments of the distribution may also matter.

Stylized use case: Optimal sowing date and variety combinations

We use CSISA-KVK trial data to demonstrate the approach. The agronomic trials cover a 5-year period in 8 districts in the Indian state of Bihar.

Input data requirements

This is an outcome-based risk assessment requiring yield or profit data for multiple years for the same site.

Toolkit workflow

Mean-Variance optimization requires only the outcome variable for multiple realizations and portfolio choices. We then use quadratic optimization to identify the frontier and optimal weights indicating the amount of land or resources that should be devoted to particular choices. The workflow is as shown in figure 12.

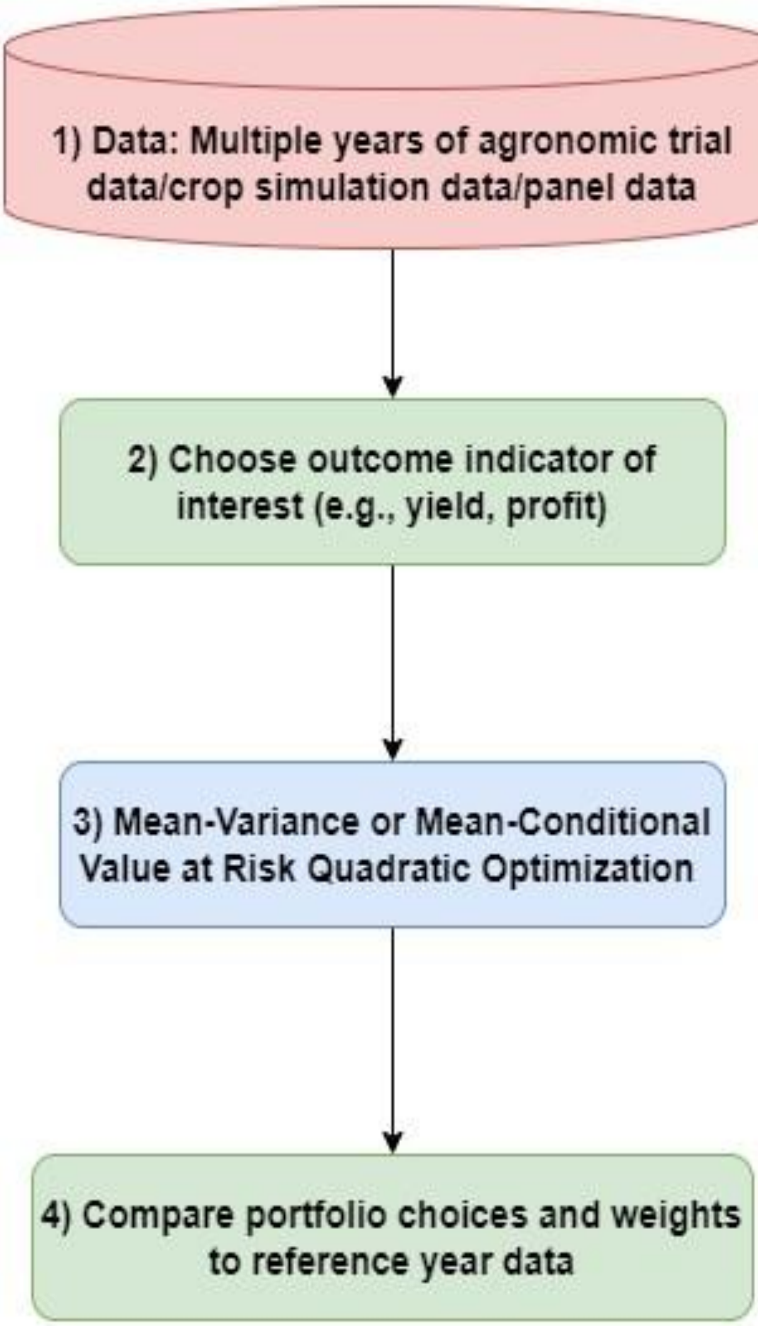

Figure 12: Mean-variance or mean-conditional value at risk (CVaR) workflow

Box 5 shows the mathematical basics for the quadratic optimization problem for mean-variance analysis and conditional value at risk (CVaR) optimization models.

Box 5: Mathematical basis for modern portfolio theory optimization models

Mean-variance optimization is solved as a quadratic programming problem:

$$\min_{\alpha_1,\ldots,\alpha_n} \sum_i \sum_j \alpha_i \alpha_j \sigma_{ij}$$

Subject to:

$$\sum_i \alpha_i y_i = \lambda$$

$$\sum_i \alpha_i = 1$$

$$\alpha_i \geq 0 \; for \; all \; i$$

where σ_{ij} is covariance of yields (or net returns) for the treatments i and j, y_i is mean yield of treatment i, α: share of land that should be allocated to the treatment and λ is the target yield. The constraint equation $\sum_i \alpha_i = 1$ restricts the shares of land allocated to technology to add up to 1.

Conditional value at risk optimization model is solved as a quadratic programming problem:

$$\min_{\alpha_1,\ldots,\alpha_n} ES_\beta = -E\left[\sum_i \alpha_i y_i | \sum_i \alpha_i y_i \leq q\right]$$

Subject to:

$$\sum_i \alpha_i y_i = \lambda$$

$$\sum_i \alpha_i = 1$$

$$\alpha_i \geq 0 \; for \; all \; i$$

where ES_β is the expected short fall, q is the βth percentile of the yield distribution ($\Pr(Y \leq q) = \beta$).

See Nayak et al (2024) for an agronomic application of the approach

Stylized outputs

Using a state level E-V optimization model of wheat yields, we find that HD-2967 sown before 10th November gives the highest returns and a risk neutral farmer would find it most beneficial. For similar results using the conditional value at risk optimization model, the interested reader can consult the replication materials.

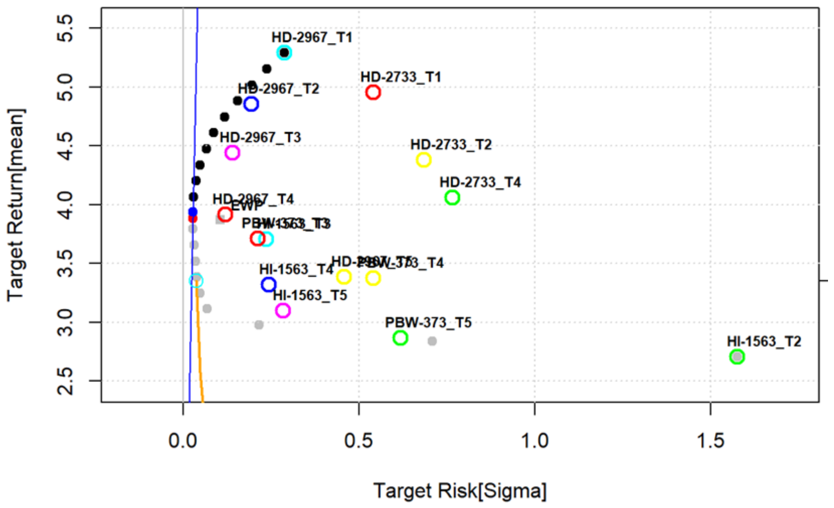

Figure 13: Planting date-varietal yield frontier for Bihar, 2016-2021.

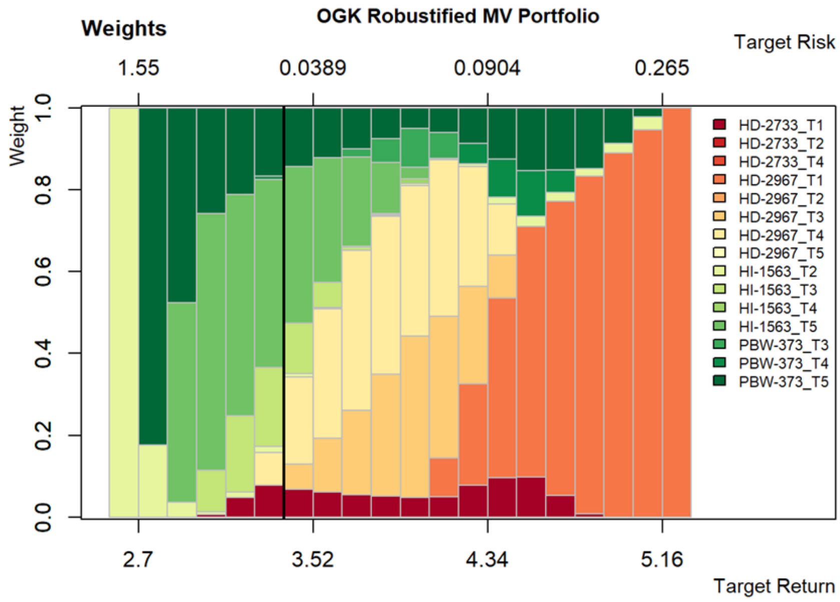

Figure 14: Optimal weights for planting date-variety yield frontier, 2016-2021.

Replication materials

https://eia2030-ex-ante.github.io/Risk_modern_portfolio_theory_EV_model/

Key references:

Nayak, H.S., Mkondiwa, M., Patra, K., Sarkar, A., Reddy, K.S., Kumar, P., Bharadwaj, S., Singh, R., and Parihar, C.M. 2025. “Risk-return trade-offs in diversified cropping systems under conservation agriculture: Evidence from a 14-year long-term experiment in north-western India”. *European Journal of Agronomy* 162: 127436. Doi: https://doi.org/10.1016/j.eja.2024.127436.

4.2. Willingness to pay bounds for second order stochastic dominance approach

Purpose

The commonly used risk measures focus on central moments (e.g., variance, conditional value at risk, skewness) of the distribution. Yield distributions over time and space are however more complicated such that one may need to consider the whole distribution when evaluating which agronomic practice will likely work where and when. The use of stochastic dominance especially second order stochastic dominance allows the relationship between the cumulative distribution function of the outcome and the expected utility maximization behavior under risk aversion. A computational approach developed by Hurley et al (2018) and Mkondiwa and Urfels (2024) allows one to compute willingness to pay lower and upper bounds for a new technology to second order stochastically dominate an old practice such that any risk averse farmer will choose the new technology.

Advantages

- Unlike mean-variance optimization, this optimization strategy considers distributional comparisons.

Disadvantages

- Computationally expensive, especially when implemented across a large area of interest.
- The comparisons are pairwise thereby requiring many combinations to come up with the best alternative for each pixel.
- Difficult to apply with survey or agronomic datasets, as it requires long time series. However, it is possible to implement the approach with Monte-Carlo simulated survey or agronomic trial datasets.

Stylized use case: Where to target sowing date advisories?

We use gridded crop growth simulation model results to identify scenarios that would be agronomically and economically beneficial even for a risk averse farmer.

Input data requirements

For the spatial ex-ante (economic) component of the model, one only needs gridded crop simulation results for each of the scenarios.

Toolkit workflow

Figure 15 shows the workflow for implementing the computational second order stochastic dominance analysis. Box 6 shows the mathematical basics underpinning the golden section search algorithm for computing the willingness to pay bounds for second order stochastic dominance.

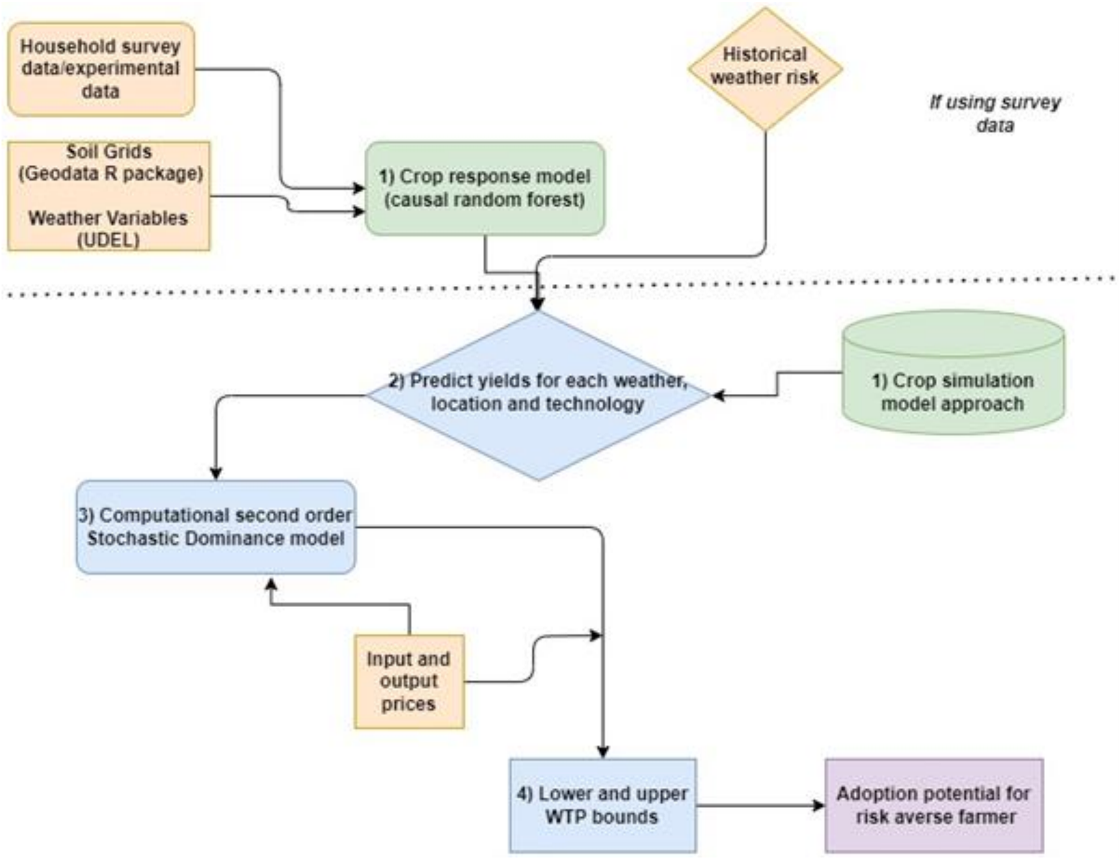

Figure 15: Risk optimization using second order stochastic dominance

Box 6: Mathematical basis for second order stochastic dominance golden section search models

Theorem relating second order stochastic dominance to risk aversion

Theorem [Meyer (1977, theorem 2)]: For cumulative distribution functions $F(x)$ and $G(x)$, $\int_0^y [G(x) - F(x)]dk(x) \geq 0,\ \forall y \in [0,1]$ (i.e., F second order stochastically dominate G)

If and only if

$$\int_0^1 u(x)dF(x) \geq \int_0^1 u(x)dG(x)\,, \forall u(x) \in U\left(\frac{k''(x)}{k(x)}, \infty\right)$$

Computing WTP bounds and benefits for risk averse farmer

We follow closely the notation and derivation by Hurley et al (2018). Consider the following notation

y: Bounded random yield where $y^L \leq y \leq y^U$,
$f^r(y)$: Yield density functions for baseline farmer practice
$f^g(y)$: Yield density functions for new agronomic innovation

$F^r(y)$: Cumulative distribution functions for baseline farmer practice
F^g: Cumulative distribution functions for new agronomic innovation

A farmer has a thrice differentiable, risk averse utility of yield function such that $U'(y) > 0$, and $U''(y) < 0$

A farmer is expected to weakly prefer the new agronomic management innovation if

$$\int_{y^L}^{y^U} U(y) f^g(y) dy \geq \int_{y^L}^{y^U} U(y)\, f^r(y) dy$$

One may compare the area under the cumulative distributions.
If

$\int_{y^L}^{y^U} F^g(z)dz \geq \int_{y^L}^{y^U} F^r(z)dz, \forall y$

Then the new agricultural innovation will be weakly preferred.
With price risks and production costs,

$$\int_{\epsilon}\int_{p>0}\int_{y^L}^{y^U} U(py - c^g)A) f^g(y|\epsilon)dy\, h(p|\epsilon)dp\phi(\epsilon)$$
$$\geq \int_{\epsilon}\int_{p>0}\int_{y^L}^{y^U} U(py - c^r)A) f^r(y|\epsilon)dy\, h(p|\epsilon)dp\phi(\epsilon)$$

WTP question is: How much wheat/rice per hectare would a risk-averse farmer be willing to give up/pay to use the new agronomic innovation?
Answer: It is the w that satisfies,

$$\int_{\epsilon}\int_{p>0}\int_{y^L}^{y^U} U(p(y - w) - c^g)A) f^g(y|\epsilon)dy\, h(p|\epsilon)dp\phi(\epsilon)$$
$$= \int_{\epsilon}\int_{p>0}\int_{y^L}^{y^U} U(py - c^r)A) f^r(y|\epsilon)dy\, h(p|\epsilon)dp\phi(\epsilon)$$

According to Hurley et al (2018), the lower WTP bound that makes any risk-averse farmer prefer new technology (in this case scenarios other than the baseline) can be derived using second order stochastic dominance as follows:

$$WTP^{LB} = \max_{w}\left\{w: \int_{y^L}^{y} F^g(z|\epsilon)dz \geq \int_{y^L-w}^{y-w} F^r(z|\epsilon)dz\ , \forall y \geq \min\{y^L - w, y^L\} \text{ and } \epsilon\right\}$$

Where WTP^{LB} is the lower bound for the willingness to pay.
Similarly, for the upper bound,

$$WTP^{UB} = \min_{w}\left\{w: \int_{y^L-w}^{y-w} F^g(z|\epsilon)dz \geq \int_{y^L}^{y} F^r(z|\epsilon)dz\ , \forall y \geq \min\{y^L - w, y^L\} \text{ and } \epsilon\right\}$$

If both lower bound and upper bound are positive, then any risk averse farmer will prefer g to r. Conversely, if **both lower bound and upper bound are negative**, then any risk averse farmer will prefer r to g.

The solution is found using golden-section search algorithm.

See Hurley et al (2018) for details.

Stylized outputs

For the sowing date application, figures 16 and 17 shows the preferred sowing date, irrigation and variety combination for even an extremely risk-averse farmer across the area of interest (Indo-Gangetic Plains). For more details on the interpretation, interested readers can consult Mkondiwa and Urfels (2024).

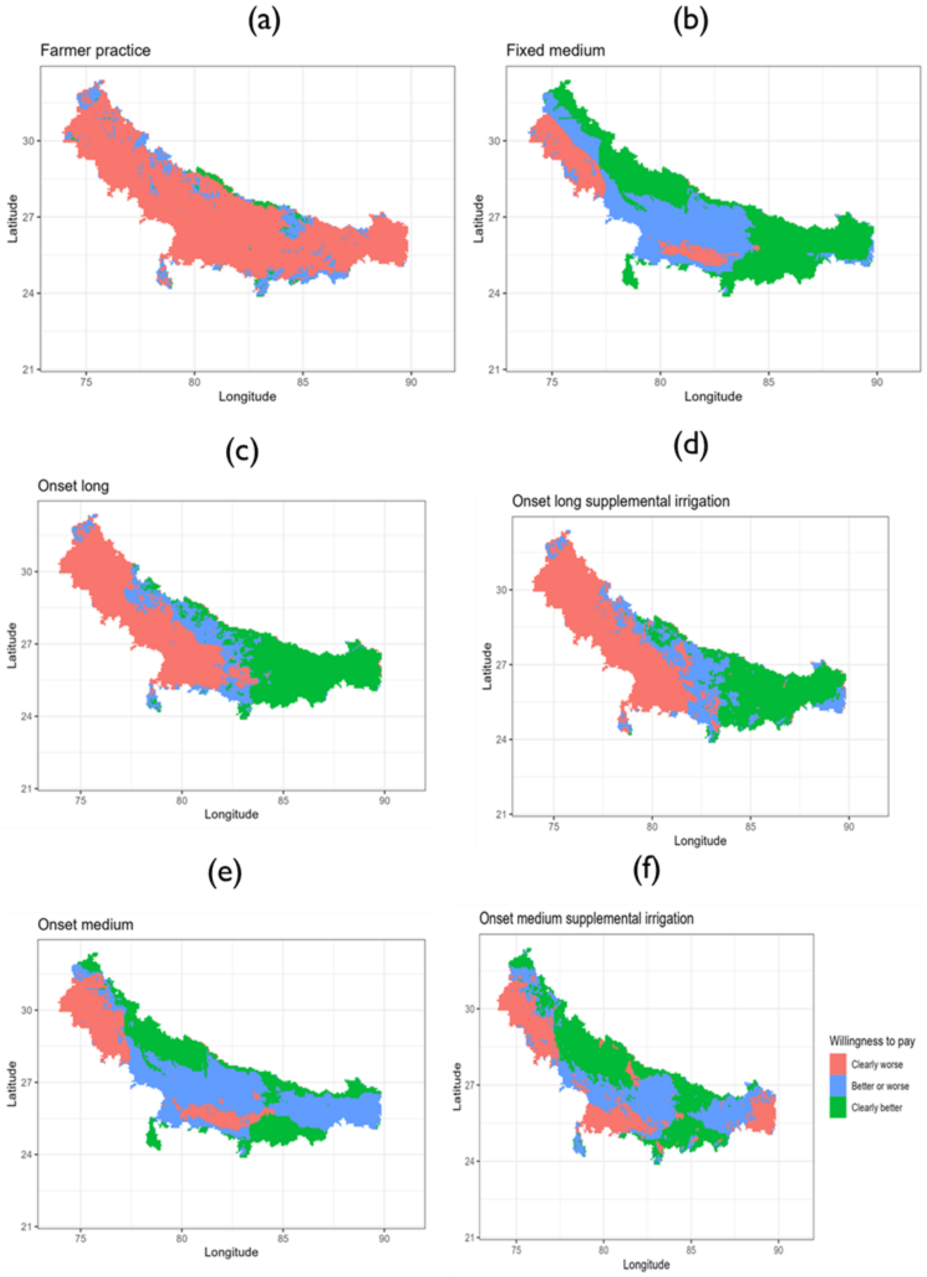

Figure 16: Willingness to pay based on partial profits for planting date scenario in comparison to fixed date with long duration rice variety strategy

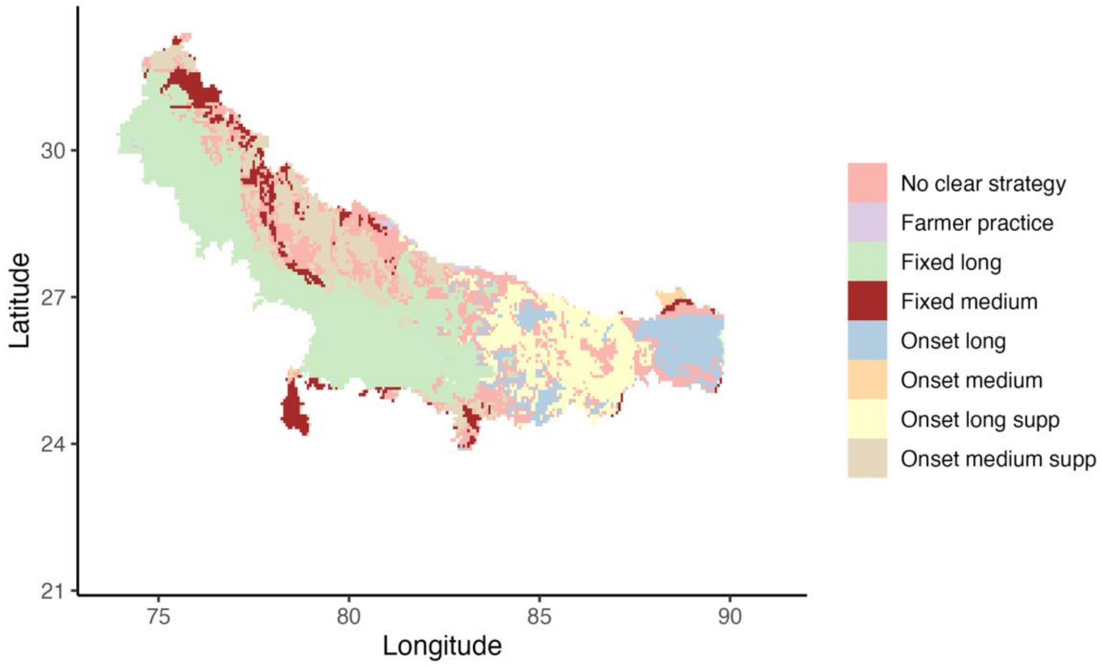

Figure 17: Robust and optimal rice planting date strategy

Replication materials

https://eia2030-ex-ante.github.io/WTP_Bounds_SOSD_Risk_Model/

Key references: For more methodological details of the approach, readers are referred to Hurley et al (2018) and Mkondiwa and Urfels (2024).

Hurley, T., Koo, J., and Tesfaye, K. 2018. "Weather risk: how does it change the yield benefits of nitrogen fertilizer and improved maize varieties in sub-Saharan Africa?" *Agricultural Economics* 49: 711-723. Doi: 10.1111/agec.12454.

Meyer, J., 1977. Second degree stochastic dominance with respect to a function. Int. Econ. Rev. 18 (2), 477–487. https://EconPapers.repec.org/RePEc:ier:iecrev:v:18:y: 1977:i:2:p:477-87.

Mkondiwa, M., and Urfels, A. 2024. "Risk-based evaluations of competing agronomic climate adaptation strategies: The case of rice planting strategies in the indo-Gangetic Plains". Agricultural Systems 2018: 104014. Doi: https://doi.org/10.1016/j.agsy.2024.104014.

5. Spatial economic surplus and return on investments (ROI) toolkit

5.1. Discounted cash flow (DCF) economic surplus framework [Incomplete]

Purpose

Economic surplus framework (Alston et al 1995) is the most used approach in agricultural economics to evaluate agricultural research and development benefits. In recent years, scholars have also suggested the use of real options approach which then helps in valuing the time to wait and dynamic complexities appropriately.

Advantages

- Relies on economic theory especially demand and supply as well as welfare economics

Disadvantages

- Given data requirements (e.g., on elasticities), the analyses are often done at aggregate level (e.g., country level). However, recent applications are exploring disaggregated analysis.

Stylized use case: Would herbicide application pay? Where?

We estimate the returns on investments to herbicide application using the economic surplus approach as a case study on how to use the approach.

Input data requirements

The economic surplus analysis approach requires data on area, yield, and prices. It also requires data on adoption profile (e.g., percent of area under the new technology), the supply and demand elasticities, yield gain and input cost changes due to the technology.

Toolkit workflow

Figure 18 shows the spatial ex ante economic surplus framework for prioritizing agronomic innovations. It rests on integrating prior adoption survey data (or studies) and current survey data to forecast the expected trends in adoption of the technology under different ecological, social and economic conditions. The adoption curves are usually fitted using S shaped functions like logistic function (following Griliches 1957), Gompertz and other similar functions. The estimates of these models are then used to predict adoption rates which are then combined using an equilibrium displacement model of linear demands and supply with administrative data on areas under the crop, the causal ML heterogeneous estimates of expected yield gains and cost reductions, and demand and supply elasticity elasticities (usually from literature).

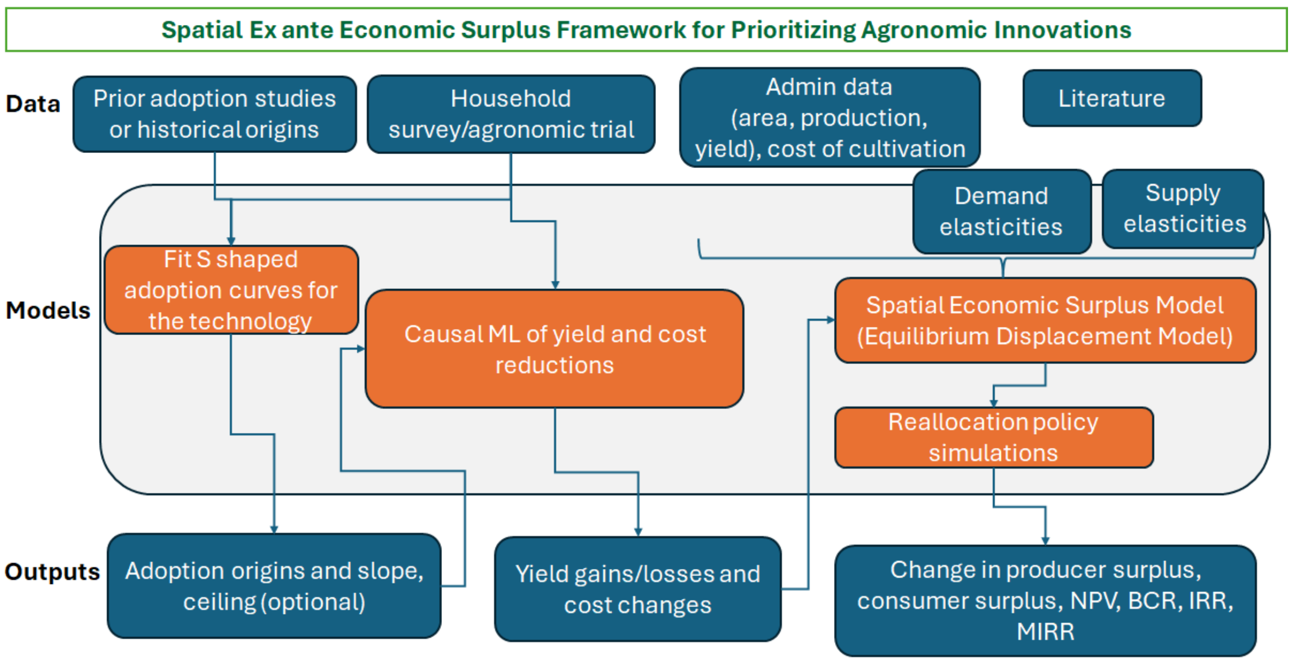

Figure 18: Workflow for spatial economic surplus model

The key outputs from this modeling are expected changes (each year in the projection) in producer, consumer and total economic surplus (approximating welfare) due to the agronomic innovation. To compare interventions across the time span, several economic measures are usually used. These include net present value, benefit-cost ratio, internal rate of return, and modified internal rate of return. A discussion of these methods for interested readers is in Hurley et al (2024).

Box 7 shows the mathematical basics of the economic surplus model. The approach rests estimating what is called a k supply shift parameter which basically shows total cost reductions/or supply shift due to the technology.

Box 7: Mathematical basis for economic surplus models

The main indicators in the economic surplus models are the supply shift parameter (K_t) due to the technology, producer surplus (PS), consumer surplus (CS), and total economic surplus ((TS), (Alston et al 1995, p.217, p.227).

K-supply shift

$$K_t = a_t\left(\frac{\frac{\tau_y}{Y}}{\epsilon} - \frac{\frac{\tau_c}{C}}{1+\frac{\tau_y}{Y}}\right)$$

Small closed economy

$$\Delta PS = P_0Q_0(K - Z)(1 + 0.5Z\eta)$$

$$\Delta CS = P_0Q_0Z(1 + 0.5Z\eta)$$
$$\Delta TS = \Delta PS + \Delta CS = P_0Q_0K(1 + 0.5Z\eta)$$

Small open economy

$$\Delta PS = \Delta TS = P_wQ_0K(1 + 0.5K\epsilon)$$

where
a_t: adoption rate of the technology
τ_y: yield gain to technology
Y: Baseline yield
τ_c: Cost change due to technology
C: Baseline costs
P_0: Baseline output price
η : elasticity of demand
ϵ: elasticity of supply
Z: $\frac{P_0 - P}{P_0}$

See Alston et al (1995) for details.

Stylized results

In terms of economic gains per hectare, we find that farmers in Munger and Banka are expected to get economic gains to herbicide weed management of above INR 6,000 per ha (figure 19).

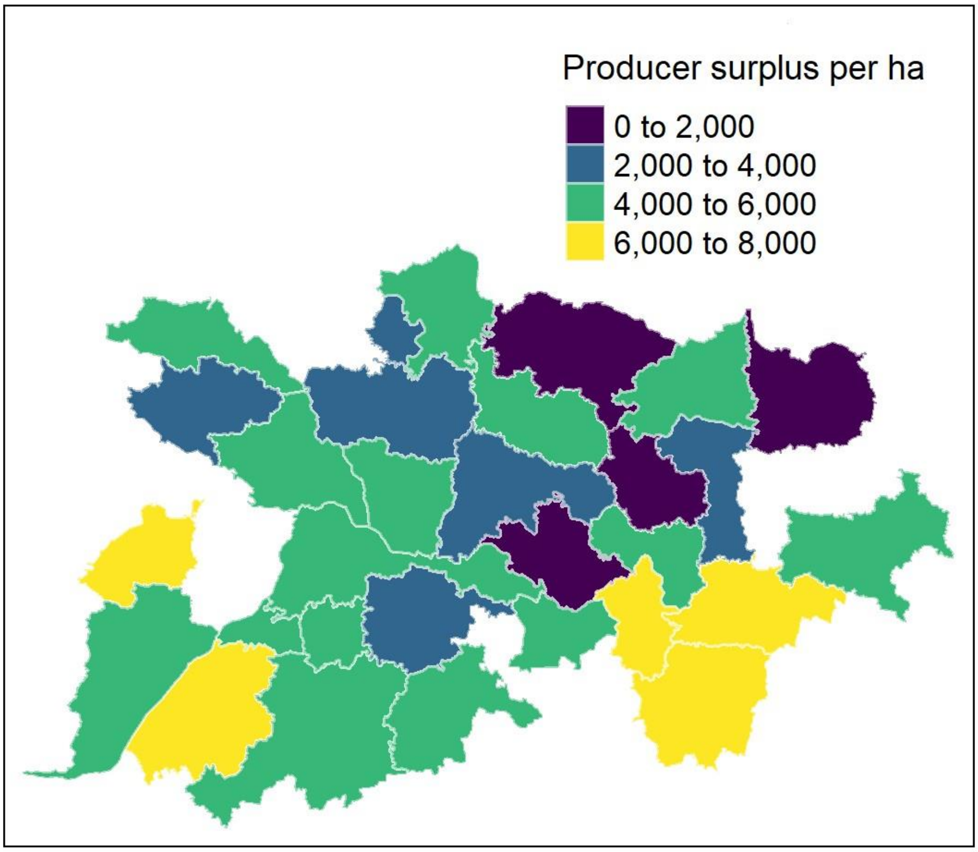

Figure 19: Producer surplus per ha

The picture is however different when we consider the aggregate economic returns at the district level (figure 20). Here, we find that even Munger and Banka generate less economic gains to the state for herbicide management because smaller total areas under wheat. A cautious approach to interpreting the results is therefore encouraged and the stakeholder preferences need to be examined before prescribing agronomic interventions.

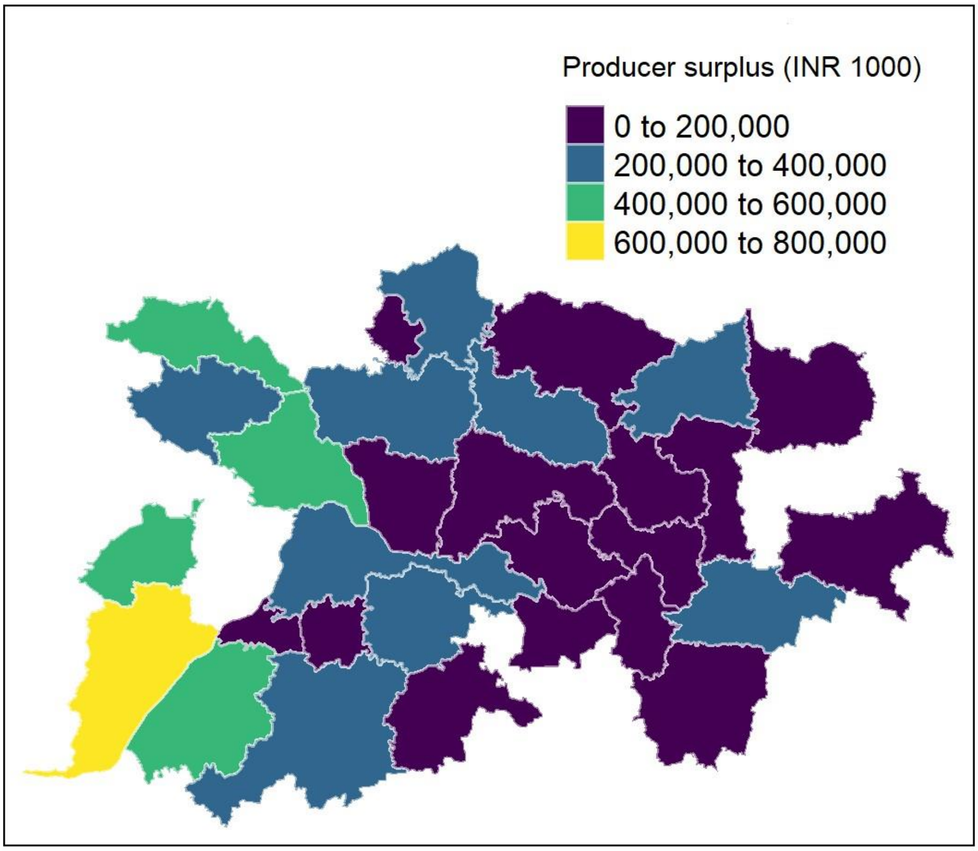

Figure 20: Annual present value producer surplus to herbicide application (2020-2030) in wheat farms in Bihar

Replication materials

https://github.com/MaxwellMkondiwa/spatial-economic-surplus

Key references

Alston, J.M., Norton, G.W., and Pardey, P. 1995. "Science under scarcity: Principles and practice for agricultural research evaluation and priority setting". CAB International. See Chapter 7 (pp. 463-498).

Hurley, T.M., Rao, X., and Pardey, P.G. 2014. "Re-examining the reported rates of return to food and agricultural research and development". American Journal of Agricultural Economics 96 (5): 1492-1504. Doi: https://doi.org/10.1093/ajae/aau047.

Mills, B.F. 1997. "Ex-ante agricultural research evaluation with site specific technology generation: the case of sorghum in Kenya". *Agricultural Economics* 16: 125-138. Doi: https://doi.org/10.1016/S0169-5150(96)01218-2.

6. Structural differentiated agronomy toolkit

6.1. Targeting research, adoption and impacts of new innovations

Purpose

A differentiated farming system approach focuses on identifying characteristics of the farming systems which farmers would find attractive. The goal is then to estimate the willingness of the farmers to pay for these farming system characteristics. The approach is an adaptation of the differentiated demand analysis approach (Nevo 2001) commonly used in new industrial organization literature.

Advantages

- It allows one to predict the adoption of new agronomic innovations, especially the ones that involve a small adjustment to existing practices. For example, adding a new modification in conservation agriculture or adding a trait to a variety.
- It helps to assess the welfare impacts of many varieties of technology unlike the traditional approach which treats them as separate inputs or goods.
- The analysis can be done at market level using aggregate product sales and price data. This is the key power of this approach in that one can do demand analysis differentiated for household level characteristics while using other key data at aggregate or market level.
- It is easy to include market power considerations for the supply sector.

Disadvantages

- Attribute data for most agricultural innovations are either difficult to define or difficult to collect. For example, while it is easy to define attributes for varieties (e.g., potential yield, disease resistance, and drought tolerance), it is difficult to ascertain these for zero tillage, fertilizers, sowing dates, and other agronomic management practices.

Stylized use case: Where to target sowing date advisories?

We use the differentiated farming systems approach by analyzing the segmentation of sowing dates and drivers associated with farmers choice of which planting dates to sow their wheat.

Input data requirements

The data required includes characteristics of the existing technology brands including prices (e.g., variety attributes) and adoption data for each of the brands (e.g., each variety).

Toolkit workflow

Figure 21 shows the framework for using differentiated agronomy framework for prioritizing agronomic innovations. It involves collecting technology characteristics/traits (e.g., from breeders for varieties) which are then combined with the brand specific adoption data from household surveys. In cases where household level data but aggregated sales data are available, one can use such data without loss of generality.

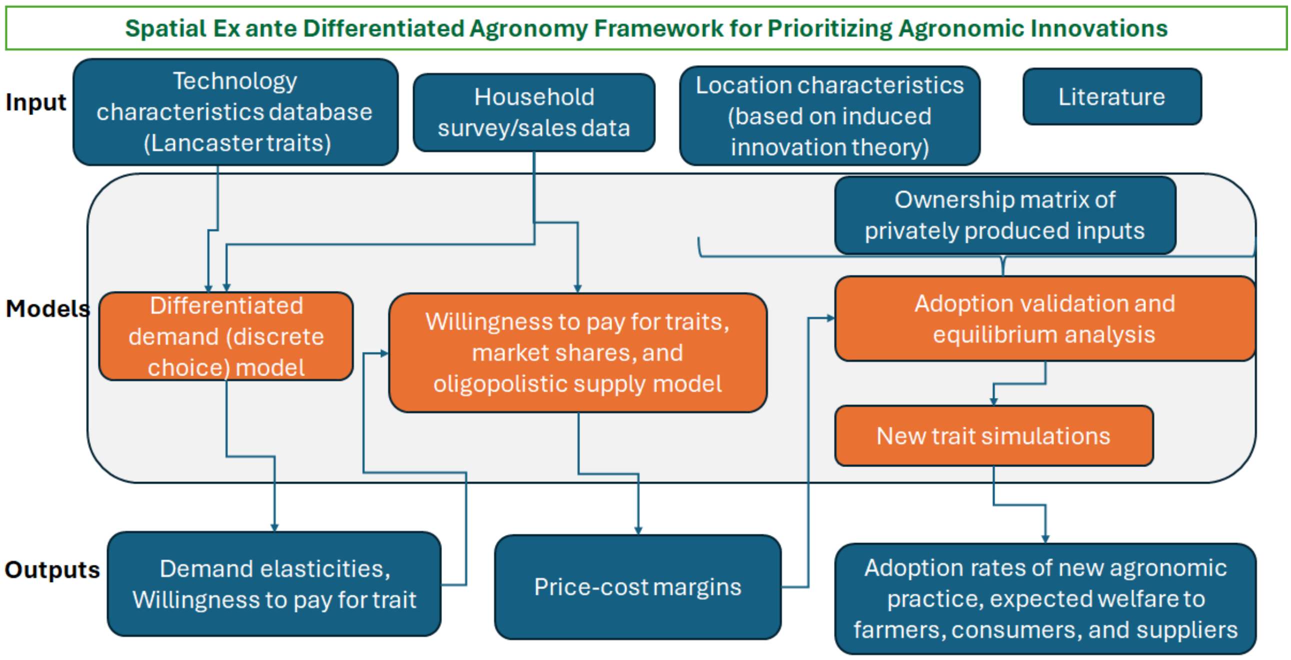

Figure 21: Workflow for differentiated agronomy targeting toolkit

Box 8 shows the mathematical basics for a differentiated agronomy model which involves estimating a linearized logit model of the adoption shares of each practice (e.g., specific brand of herbicide) as explained by prices of each brand, characteristics of the brand (e.g., herbicide effective for broadleaf/grassy weeds, e.t.c) and household farmer characteristics.

Box 8: Mathematical basics for differentiated agronomy models

The differentiated agronomy model uses the differentiated demand model (see Nevo 2001, Ciliberto et al 2019, Mkondiwa 2019) and assumptions about an oligopolistic supply sector to predict adoption of new practices embodying particular combination of traits.

Differentiated demand model

$$log\left(\frac{S_j}{S_0}\right) = \beta_0 - \alpha p_j + \sum_k \beta_k x_k + \sum_m \eta_m z_m + \Gamma + \epsilon_{ij}$$

Where s_j is the adoption share of brand j, s_0 is the adoption share of outside brand (usually considered no adoption or adoption of old practices), p_j is the price of brand j , x_k is the vector of characteristics for each of the brands, z_m is a vector of farm and farmer characteristics.Γ are the unobserved characteristics.

The estimated model (ideally applying quasi-experimental approaches,e.g., instrumental variables estimator) allows one to compute willingness to pay for each of the traits as follows:

$$WTP_k = -\frac{\beta_k}{\alpha}$$

Oligopolistic supply module

Given the demand parameters, one can assume profit maximizing oligopolistic supply firm sector with an ownership structure (Ω) to estimate the price cost margin (p-mc) which is a sufficient statistic for the farmer losses due to market structure.

$$\Pi_f = \sum_{j \in \mathcal{F}_f} (p_j - mc_j) M s_j(p) - C_f$$

$$p - mc = \Omega^{-1} s(p)$$

For details: See Nevo (2001), Ciliberto et al (2019) and Mkondiwa (2019)

Stylized outputs

Figure 22 shows an example of a stylized output from a demand model assessing the likely adoption of different sowing dates. For example, it is clear that the rice season affects sowing dates. We have not yet completed the full analytics for the application. However, in the key references section we provide a list of applications mostly based on varietal targeting.

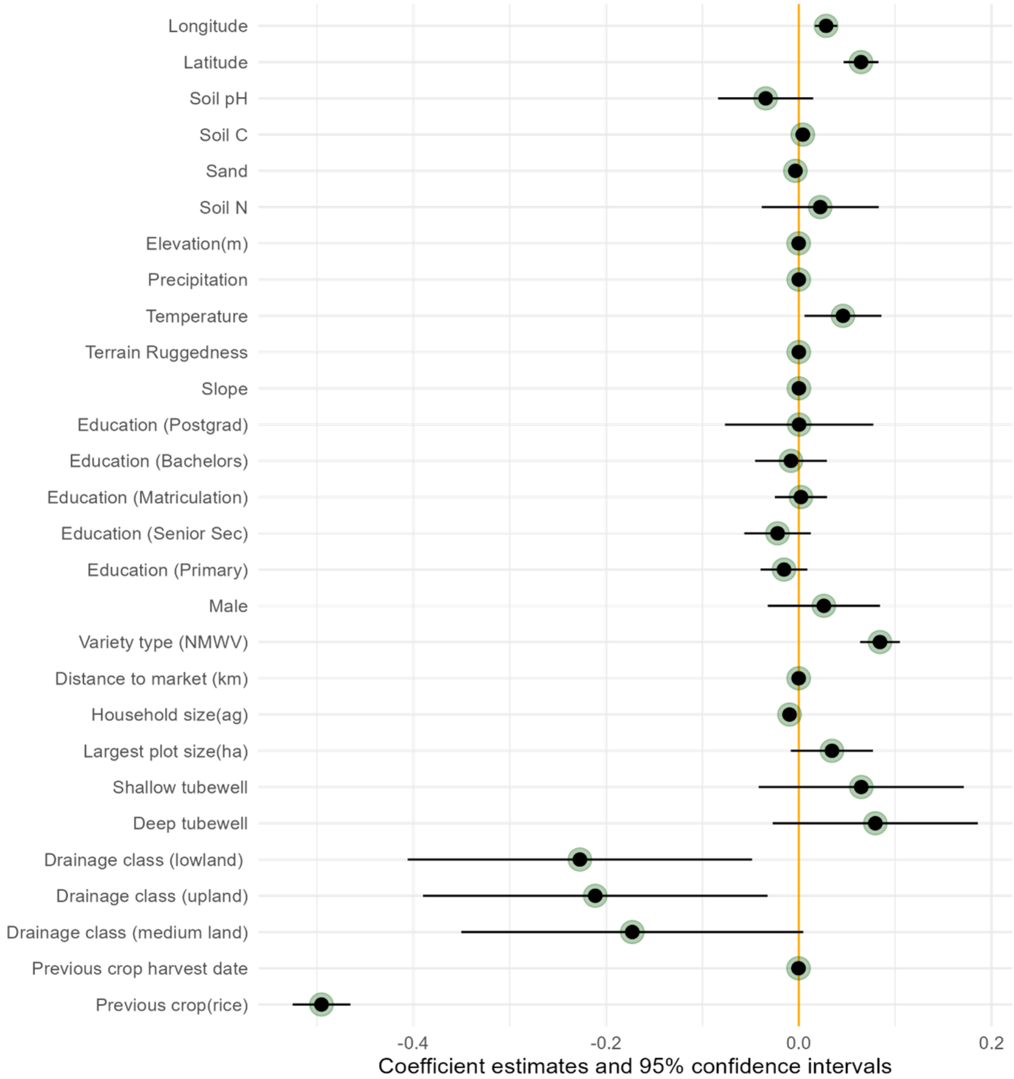

Figure 22: Factors affecting timely sowing of wheat

Replication materials

https://github.com/EiA2030-ex-ante/Differentiated_agronomy_targeting

Key references

Ciliberto, F., Moschini, G., and Perry, E.D. 2019. "Valuing product innovation: genetically engineered varieties in US corn and soybeans." *The Rand Journal of Economics* 50 (3): 615-644. Doi: https://doi.org/10.1111/1756-2171.12290.

Lee, S., and Moschini, G. 2020. "On the value of innovation and extension information: SCN-resistant soybean varieties." *American Journal of Agricultural Economics* 104 (4): 1177-1202. Doi: https://doi.org/10.1111/ajae.12283.

Mkondiwa, M. 2019. "Economics of scaling agricultural research recommendations to up-scale adoption and impact." PhD Dissertation. University of Minnesota. Url: https://hdl.handle.net/11299/206645. See Chapter 3.

Nevo, A. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307-342. Doi: https://doi.org/10.1111/1468-0262.00194.

7. Other spatial ex ante economic analysis toolkits

7.1. List of alternative workflows

The workflows we have presented in this version of the manual are limited in scope. Many other workflows are used in practice. These include more complicated methods, such as spatial micro-simulation analysis, real options analysis, spatial multicriteria decision tools (e.g., Analytic Hierarchy Process), spatial multi-objective optimization tools (e.g., heuristic models using genetic algorithms), spatial agent-based modeling, spatial mathematical programming models (e.g., positive mathematical programming and crop mix agricultural sector models), spatial dynamic programming models, spatial regression discontinuity and remote sensing applications. A common challenge with each of these complicated methods (that we did not cover in this manual) is that they require a textbook or at least several pages just to present the basics. For interested readers, we provide below key reference materials for some of these methods.

Spatial micro-simulation analysis

Spatial micro-simulation analysis is a variant of the economic surplus approach which adds Deaton (1989) net benefit ratio idea to simulate the distributional impacts of price changes due to negative (e.g., droughts) and positive (e.g., technological change) shocks. The approach and applications are well summarized in a manual by Martin and Minot (2021).

Real options analysis

Another variant of the economic surplus approach is the real options analysis which extends the computation of discounted cash flows which assume that the decision maker will make the decision at the beginning of the horizon (e.g., by calculating net present values) to cases in which there is a value of flexibility and anticipation. In cases of uncertainty (e.g., monsoon onset uncertainty), there may be an economic value to waiting to gather data and forecasts which then allows choice of a better agronomic intervention. Real options analysis is relatively new in analyzing agricultural interventions (see, Andoseh et al 2014 and Yanore et al 2023).

Spatial multicriteria decision tools

Agronomic interventions are increasingly expected not only to increase yields but also reduce risk, improve soil health, enhance resilience, maintain biodiversity and environmental sustainability of farming systems, improve the economic well-being of smallholder farmers and foster inclusive development for landowners and landless laborers. These multiple and sometimes conflicting goals require approaches to making decisions that help analyze the trade-offs. Several spatially differentially multi-criteria decision tools have been used in practice. The non-spatial variants include Analytic Hierarchy Process (AHP) (Hartwick and Jenssen 2000; De Marinis and Sali 2020) and Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) (Gbegbelegbe et al 2024).

Spatial mathematical programming models

Spatial mathematical programming models (also known as price endogenous agricultural sector mathematical programming models) are multi-region, multi-crop and multi-input mathematical programming models. These models have the advantage of being comprehensive in capturing the effect of multiple agronomic changes. As with any other mathematical programming model though, these can be prone to overspecialization of land use and land management recommendations—an undesirable feature. Recent applications therefore use either positive mathematical programming (i.e., adding cost function constraints following Howitt 1995) or crop mix approach (i.e., restricting solutions to historical crop mixes following McCarl 1982). For reviews of recent applications of each of these approaches, interested readers can read Fei & McCarl (2023) and Mkondiwa & Apland (2022) for crop mix applications, and Merel & Howitt (2014) for positive mathematical programming applications.

Spatial agent-based modeling

Spatial agent-based modeling (ABM) involves analyzing individual decisions and complex interactions among individuals. The advantage of these models is that one can incorporate multiple stressors, multiple goals, and multiple sectors while still analyzing individual behaviors (e.g., cognitive biases). Recent ABM applications in evaluating agronomic innovations include Huber et al (2022) and Huber et al (2024) who developed an ex-ante agent-based model called FARMIND (FARM interaction and Decision-making) which allows integration of behavioral economics in spatial mathematical programming models thereby allowing evaluation of nudge interventions.

Spatial Bayesian optimization

Bayesian optimization is an approach to optimize an arbitrary objective function and selecting the next samples to improve the value from that objective function. This is essentially what agronomic field experiments try to accomplish. As a spatial ex ante economic analysis, it can be used to combine a machine learning algorithm through a gaussian process model and upper confidence bounds to develop recommendations for choosing the best agronomic strategies while selecting locations for which the testing could happen to generate risk robust recommendations. The workflow can be represented using figure 21 below.

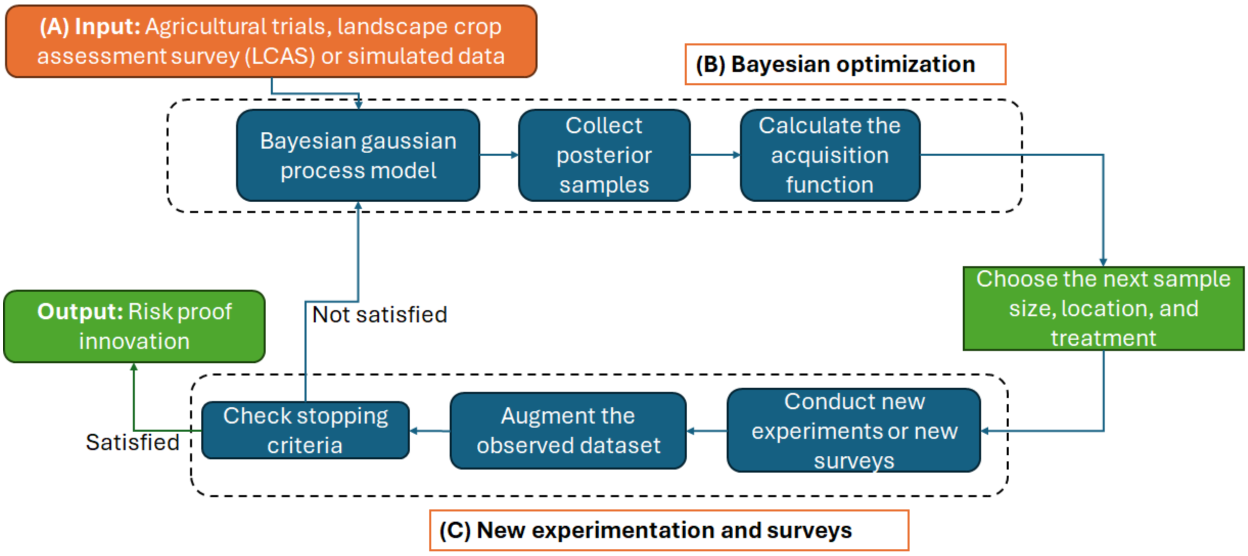

Figure 21: Bayesian optimization framework for discovering risk proof agricultural innovations (adapted from Lei et al 2021).

The Bayesian optimization approach has been extensively used in drug discovery, plant breeding (Diot and Iwata 2022) and recently in ex ante analysis of site-specific nitrogen management recommendations (Saikai et al 2020).

7.2. Five cultures of spatial ex ante economic analysis of agronomy interventions

Breiman (2001) called the causal inference and prediction tools commonly used in statistics and computer science as the "two cultures" of statistical analysis. In recent years, a growing literature has proposed a third culture called causal machine learning which integrates the best of Breiman's two cultures. Parallel to the causal machine learning paradigm is the knowledge guided machine learning (KGML) (e.g., Liu et al 2022) which integrates mechanistic models and machine learning models. These three cultures, however, just provide an assessment of the technological gains but do not provide decisions on what is best for that farmer or for society. Decision support tools and recommender systems are usually then used to guide farmers in taking the best choices. These can be simple like yield maximization, profit maximization, and risk minimization. Finally, there are tools that consider the complexity of the systems, and the interrelatedness of farmer decisions, their environment, and trade mostly studied using system optimization models.

We can consider the "cultures" as the principles needed for a new agronomy that is guided by spatial ex ante economic analytical tools. In that respect, we propose a five cultures framework that ensures that agronomy interventions recommended to farmers (i) have clear causal impacts (causal inference), (ii) are predictive to wider landscape scales beyond the plot or farm (predictive), (iii) are relevant to individual farmer or farm or plot contexts and goals (causal AI), (iv) are user friendly, inclusive, and take into account the inequality implications, and (v) have clear benefits taking into account equilibrium feedbacks. We argue that this framework (figure 24) can be implemented in one integrated five cultures workflow thereby carefully incorporating the demands associated with changing the livelihoods of poor smallholder farmers using the toolkits presented in the manual.

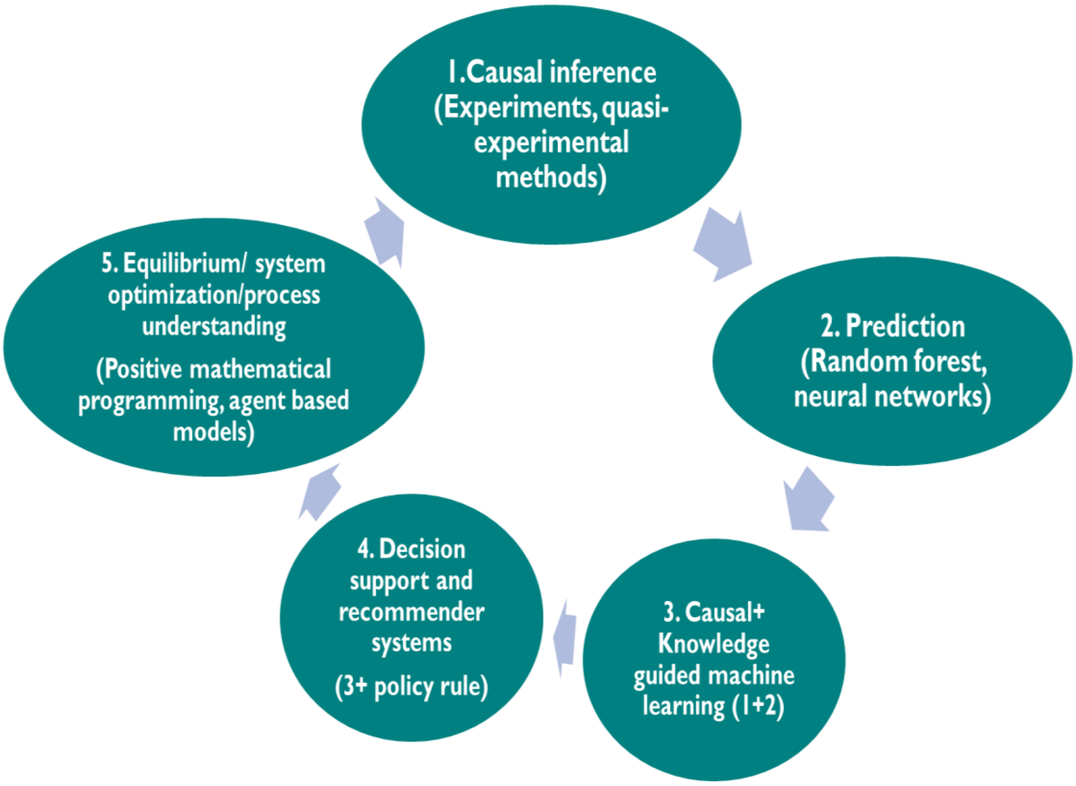

Figure 24: The five cultures in spatial ex ante economic analysis of agronomy interventions. Source: Mkondiwa et al (2023).

8. Conclusion

This manual has summarized toolkits for conducting spatial ex ante analyses using a case study of sowing date advisories. The workflows can be used for any other agricultural innovations including fertilizers, varieties, weed management practices, and irrigation practices. These toolkits are selected to showcase reproducible workflows that can be applied to other agronomic practices and in other areas of interest. In on-going research, we are also using mathematical optimization approaches, agent-based modelling approaches, knowledge-guided machine learning and spatial multi-criteria approaches for targeting and prioritizing agricultural innovations. We believe automating research around these methods will allow agronomists and other stakeholders to quickly make scientific discoveries and respond to ever-changing agronomic environments due to climate variability and change, and market conditions.

The manual will be updated annually to incorporate advances in the methods and make corrections to any mistakes in the current version. Readers are encouraged to report to us if they find any errors (expected to be many!) in the current version.

Acknowledgements

We thank the OneCGIAR Initiative on Excellence in Agronomy (INV-005431) and the Cereal Systems Initiative for South Asia (CSISA) for funding this work. This version of the draft manual has not been reviewed thoroughly and should be used cautiously. All errors are ours.

Data and code availability

The data and R scripts are available in the github repository: https://github.com/EiA2030-ex-ante/spatial_exante_sop_book.

References

Alston, J.M., Norton, G.W., and Pardey, P. 1995. "Science under scarcity: Principles and practice for agricultural research evaluation and priority setting". CAB International. See Chapter 7 (pp. 463-498).

Andoseh, S., Bahn, R., and Gu, J. 2014. "The case for a real options approach to ex-ante cost-benefit analysis of agricultural research projects". Food Policy 44: 2018-226. Doi: https://doi.org/10.1016/j.foodpol.2013.09.016.

Antle, J.M. 1983. "Testing the stochastic structure of production: A flexible moment-based approach". Journal of Business and Economic Statistics 1(3): 192-201. Doi: 10.1080/07350015.1983.10509339.

Antle, J.M. 2010. "Asymmetry, partial moments and production risk." American Journal of Agricultural Economics 92(5): . Doi: https://doi.org/10.1093/ajae/aaq077.

Athey, S., Tibshirani, J., and Wager, S. 2019. "Generalized Random Forests." The Annals of Statistics 47(2): 1148-1178. Doi: 10.1214/18-AOS1709.

Athey, S., and Wager, S. 2021. “Policy learning with observational data”. Econometrica. Url: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15732.

Breiman, L.2001. “Statistical Modelling: The Two Cultures”. Statistical Science 16(3): 199-231. Doi: https://doi.org/10.1214/ss/1009213726.

Binswanger, H.P. 1986. “Evaluating Research System Performance and Targeting Research in Land-abundant Areas of Sub-Saharan Africa.” World Development 14(4): 469-475. Doi: https://doi.org/10.1016/0305-750X(86)90063-X.

Ciliberto, F., Moschini, G., and Perry, E.D. 2019. “Valuing product innovation: genetically engineered varieties in US corn and soybeans.” The Rand Journal of Economics 50 (3): 615-644. Doi: https://doi.org/10.1111/1756-2171.12290.

De Marinis, P., and Sali, G. 2020. “Participatory analytic hierarchy process for resource allocation in agricultural development projects”. Evaluation and Program Planning 80: 101793. Doi: https://doi.org/10.1016/j.evalprogplan.2020.101793.

Deaton, A. 1989. “Rice prices and income distribution in Thailand: A non-parametric analysis”. Economic Journal 99: 1-37. Doi: https://doi.org/10.2307/2234068.

Di Falco, S., Chavas, J., and Smale, M. 2007. “Farmer management of production risk on degraded lands: the role of wheat variety diversity in the Tigray region, Ethiopia.” Agricultural Economics 36: 147-156. Doi: https://doi.org/10.1111/j.1574-0862.2007.00194.x.

Di Falco, S., and Chavas, J. 2009. “On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia”. American Journal of Agricultural Economics 91(3): 599-611. Doi: https://doi.org/10.1111/j.1467-8276.2009.01265.x.

Diot, J., and Iwata, H. 2022. “Bayesian optimization for breeding schemes”. Frontiers in Plant Science 13: 1050198. Doi: https://doi.org/10.3389/fpls.2022.1050198.

Fei, C.J., and McCarl, B.A. 2023. “The role and use of mathematical programming in agricultural, natural resource, and climate change analysis”. Annual Review of Resource Economics 15: 383-406. Doi: https://doi.org/10.1146/annurev-resource-101422-041745.

Gbegbelegbe, S., Alene, A., Swamikannu, N., and Frija, A. 2024. “Multi-dimensional impact assessment for priority setting of agricultural technologies: An application of TOPSIS for the drylands of sub-Saharan Africa and South Asia”. PLOS One. Doi: https://doi.org/10.1371/journal.pone.0314007.

Goldman, A. 1993. “Agricultural Innovation in Three Areas of Kenya: Neo-Boserupian Theories and Regional Characterization.” Economic Geography 69(1): 44-71. Url: http://www.jstor.org/stable/143889.

Griliches, Z. 1957. “Hybrid corn: An exploration in the economics of technological change.” Econometrica 25 (4): 501-522. Doi: https://doi.org/10.2307/1905380.

Hartwich, F., and Janssen, W. 2000. “Setting research priorities: an example from agriculture using the Analytic Hierarchy Process”. Research evaluation 9 (3): 201-210. Doi: https://doi.org/10.3152/147154400781777250.

Huber, R., Xiong, H., Keller, K., and Finger, R. 2022. "Bridging behavioral factors and standard bio-economic modelling in an agent-based modelling framework". Journal of Agricultural Economics 73 (1): 35-63. Doi: https://doi.org/10.1111/1477-9552.12447.

Huber, R., Kreft, C., Spatl, K., and Finger, R. 2024. "Quantifying the importance of farmer' behavioral factors in ex-ante assessments of policies supporting sustainable farming practices". Ecological Economics 224: 108303. Doi: https://doi.org/10.1016/j.ecolecon.2024.108303.

Hurley, T.M., Rao, X., and Pardey, P.G. 2014. "Re-examining the reported rates of return to food and agricultural research and development". American Journal of Agricultural Economics 96 (5): 1492-1504. Doi: https://doi.org/10.1093/ajae/aau047.

Hurley, T., Koo, J., and Tesfaye, K. 2018. "Weather risk: how does it change the yield benefits of nitrogen fertilizer and improved maize varieties in sub-Saharan Africa?" Agricultural Economics 49: 711-723. Doi: 10.1111/agec.12454.

Lee, D.R., Edmeades, S., Nys, E., McDonald, A., Janssen, W. 2014. "Developing local adaptation strategies for climate change in agriculture: A priority-setting approach with application to Latin America". Global Environmental Change 29: 78-91. Doi: https://doi.org/10.1016/j.gloenvcha.2014.08.002.

Lei, B., Kirk, T.Q., Bhattarcharya, A., Pati, D., Qian, X., Arroyave, R., and Mallick, B.K. 2021. "Bayesian optimization with adaptive surrogate models for automated experimental design". NPJ Computational Materials 194. Doi: https://doi.org/10.1038/s41524-021-00662-x.

Liu, L., Xu, S., Tang, J., Guan, K., Griffis, T. J., Erickson, M. D., Frie, A. L., Jia, X., Kim, T., Miller, L. T., Peng, B., Wu, S., Yang, Y., Zhou, W., Kumar, V., and Jin, Z. 2022. "KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N2O emission using data from mesocosm experiments." Geosci. Model Dev., 15, 2839–2858, https://doi.org/10.5194/gmd-15-2839-2022.

Martin, W., and Minot, N. "Prioritizing agricultural investments across commodities for income growth and poverty reduction: Methods and applications". Technical paper 3. IFPRI-MCC Series: Prioritizing agricultural investments for income, poverty reduction and nutrition. Url: https://doi.org/10.2499/p15738coll2.134800.

McCarl, B.A. 1982. "Cropping activities in agricultural sector models: a methodological proposal". American Journal of Agricultural Economics 64 (4): 768-772. Doi: https://doi.org/10.2307/1240588.

McCullough, E.B., Quinn, J.D., Simons, A.M. 2020. "Profitability of climate-smart soil fertility investment varies widely across sub-Saharan Africa." Nature Food 3:275-285. Doi: https://doi.org/10.1038/s43016-022-00493-z.

Merel, P., and Howitt, R. 2014. "Theory and application of positive mathematical programming in agriculture and the environment". Annual Review of Resource Economics 6: 451-470. Doi: https://doi.org/10.1146/annurev-resource-100913-012447.

Mills, B.F. 1997. "Ex-ante agricultural research evaluation with site specific technology generation: the case of sorghum in Kenya". *Agricultural Economics* 16: 125-138. Doi: https://doi.org/10.1016/S0169-5150(96)01218-2.

Mkondiwa, M. 2019. “Economics of scaling agricultural research recommendations to up-scale adoption and impact.” PhD Dissertation. University of Minnesota. Url: https://hdl.handle.net/11299/206645. See Chapter 3.

Mkondiwa, M., and Apland, J. 2022. “Inter-district food flows in Malawi”. Food Security 14: 1553-1568. Doi: https://doi.org/10.1007/s12571-022-01302-y.

Mkondiwa, M., Kumar, V., Craufurd, P., McDonald, A., and Malik, R. 2023. “Empowering sustainability transitions in agri-food systems using big data and causal artificial intelligence (causal AI”. Lead lecture presented at Haryana Agricultural University International Conference on “Strategies for Global Food and Nutritional Security, Sustainability and Wellness”. Url: https://scholar.google.com/scholar?oi=bibs&cluster=3522962238549821118&btnI=1&hl=en.

Mkondiwa, M., Hurley, T.M., and Pardey, P.G. 2024. “Closing the gaps in experimental and observational crop response estimates: a bayesian approach”. Q Open 4. Doi: https://doi.org/10.1093/qopen/qoae017.

Mkondiwa, M., and Urfels, A. 2024. “Risk-based evaluations of competing agronomic climate adaptation strategies: The case of rice planting strategies in the indo-Gangetic Plains”. Agricultural Systems 2018: 104014. Doi: https://doi.org/10.1016/j.agsy.2024.104014.

Mourtzinis, S., Mitchell, P., Esker, P., Cerrudo, A., Naeve, S., and Conley, S. 2023. “Field-level yield benefits and risk effects of intensive soybean management across the US”. *Field Crops Research* 301: 109012. Doi: https://doi.org/10.1016/j.fcr.2023.109012.

Nayak, H.S., Mkondiwa, M., Patra, K., Sarkar, A., Reddy, K.S., Kumar, P., Bharadwaj, S., Singh, R., and Parihar, C.M. 2025. “Risk-return trade-offs in diversified cropping systems under conservation agriculture: Evidence from a 14-year long-term experiment in north-western India”. *European Journal of Agronomy* 162: 127436. Doi: https://doi.org/10.1016/j.eja.2024.127436.

Nevo, A. 2001. “Measuring Market Power in the Ready-to-Eat Cereal Industry.” *Econometrica* 69 (2): 307-342. Doi: https://doi.org/10.1111/1468-0262.00194.

Palmas, S., and Chamberlin, J. 2020. “Fertilizer profitability for smallholder maize farmers in Tanzania: A spatially-explicit exante analysis”. PLoS ONE 15(9): e0239149. Doi: https://doi.org/10.1371/journal.pone.0239149.’

Pardey, P.G., Andrade, R.S., Hurley, T.M., Rao, X., and Liebenberg, F.G. 2016. “Returns to food and agricultural R&D investments in Sub-Saharan Africa, 1975-2014”. *Food Policy* 65: 1-8. Doi: https://doi.org/10.1016/j.foodpol.2016.09.009.

Saikai, Y., Patel, V., and Mitchell, P.D. 2020. “Machine learning for optimizing complex site-complex management”. Computers and Electronics in Agriculture 174: 105381. Doi: https://doi.org/10.1016/j.compag.2020.105381.

Shi, G., Chavas, J., and Lauer, J. 2013. “Commercialized transgenic traits, maize productivity and yield risk”. *Nature Biotechnology* 31: 111-114. Doi: https://doi.org/10.1038/nbt.2496.

Yanore, L., Sok, J., Oude Lansink, A. 2023. “Anticipate, wait or don’t invest? The strategic net present value approach to study expansion decisions under policy uncertainty”. Agribusiness 39 (2): 535-548. Doi: https://doi.org/10.1002/agr.21780.