Spatial Probabilistic and Risk Assessment Tools for Targeting CSA Advisories

Maxwell Mkondiwa
Spatial Economist, CIMMYT, New Delhi

CIMMYT-BISA Training
Jabalpur, December 14th, 2022

Learning outcomes

1. Share ideas on how to analyse climate related risks and target CSA technologies to specific farmers and locales
2. Explore latest methods in causal machine learning (ML), spatial modelling and Bayesian statistics in analysing risky decisions like CSA advisories

- Lecture materials available here: https://github.com/MaxwellMkondiwa/Training_Materials

- Available for one on one chats (tomorrow or through a visit in future)
- Email: m.mkondiwa@cgiar.org

CIMMYT

Why spatial? Probabilistic? Risk? Targeting

- **Spatial:** Agriculture is a spatial and context specific activity-weather, geography, ecology matter.
  - 4th agricultural evolution likely in data and precision agriculture (after chemical, mechanical, biological revolutions)

- **Probabilistic:** Farmers do not have certitude

- **Risk/Uncertainty:** Possibility of adversity or loss. The state of nature is not known.

- **Targeting:** There is no average farmer. Need a way for advisories to be best for each farmer

CIMMYT

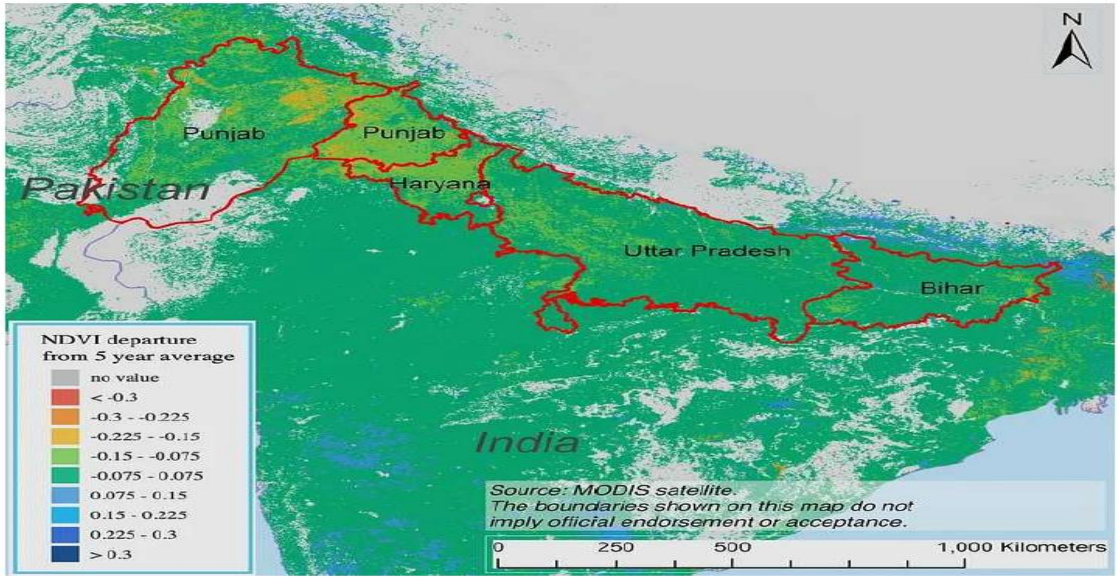
Motivation: 2021-22 Heat stress and drought shocks are costly

Wheat versus heat

Urgent action is required to mitigate effects of temperature extremes in South Asia, which threaten wheat production and human health.

By Alison Bentley

May 20, 2022

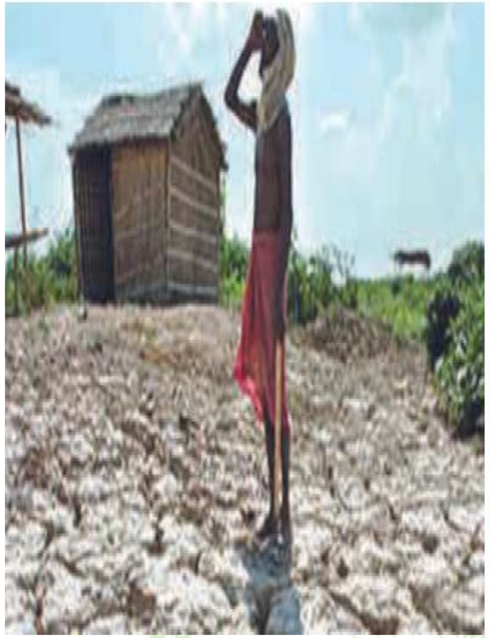


Departure of the normalized difference vegetation index (NDVI) during the period from March 22 to April 7 from the average of the previous five years. The NDVI is a measure of the leaf area and the greenness of vegetation. The yellow areas in the Punjabs of India and Pakistan, as well as in the state of Haryana, indicate that wheat matured earlier than normal due to elevated temperatures. Maximum temperatures reached 40ºC on March 15 and remained at or above this level throughout the wheat harvesting period. (Map: Urs Schulthess/CIMMYT).

Printed from
THE TIMES OF INDIA

Bihar govt OKs Rs 500-crore aid to 11 drought-hit districts

TNN | Oct 14, 2022, 04.53 AM IST



PATNA: At a time when many states are facing incessant late monsoon rainfall, the Bihar cabinet on Thursday declared 7,841 villages of 11 districts as drought hit and sanctioned Rs 500 crore in aid.
Every family in these villages across 937 panchayats in 96 blocks will receive special assistance worth Rs 3,500 each from the disaster management department, additional chief secretary (cabinet secretariat) S Siddharth said.

The disaster management department had identified Jehanabad, Gaya, Aurangabad, Sheikhpura, Nawada, Munger, Lakhisarai, Bhagalpur, Banka, Jamui and Nalanda, as the worst-hit districts.

CIMMYT

Motivation

- Agricultural production is risky. Multiple and complex shocks.
  - Agricultural production risks: weather, biotic, abiotic [Hurley et al 2010. A Review of Agricultural Production Risk in the Developing World]
  - Price risks [Boyd and Bellemare 2020. Microeconomics of Agricultural Price Risk. Annual Review of Resource Economics]
- All these change our assessment of what is beneficial, high yielding or profitable.
- Farmers are heterogenous in terms of many factors (e.g., land size, incomes, e.t.c) but also in terms of risk preferences.
- Yet much of the spatially explicit assessment of yield and profitability of technology adoption do not take these risks in consideration.

CIMMYT

Motivation

- Terminology:
  - Risk: probabilities can be estimated [Today]
  - Uncertainty: Not aware of probabilities
- Beyond production and price risks

- There are multiple risks to do with us

- Parameter risk/uncertainty: We do not propagate all the risk components

- Risks due to heterogeneity: Going beyond the average? Average farmer doesn't exist.

- I will present on-going work on developing a framework that allows a probabilistic assessment of the profitability of agronomic innovations and the adoption potential for a risk averse farmer.

My on-going work in this area

- Work in the Cereal Systems Initiative for South Asia (CSISA) project
- Planting date advisories
  - Using crop simulation data
  - Using survey data
- Return on investment analyses for CSA technologies
- Participatory Climate Adaptation Tool under Excellence in Agronomy (EiA)
- Bihar Drought Impact Survey

Outline

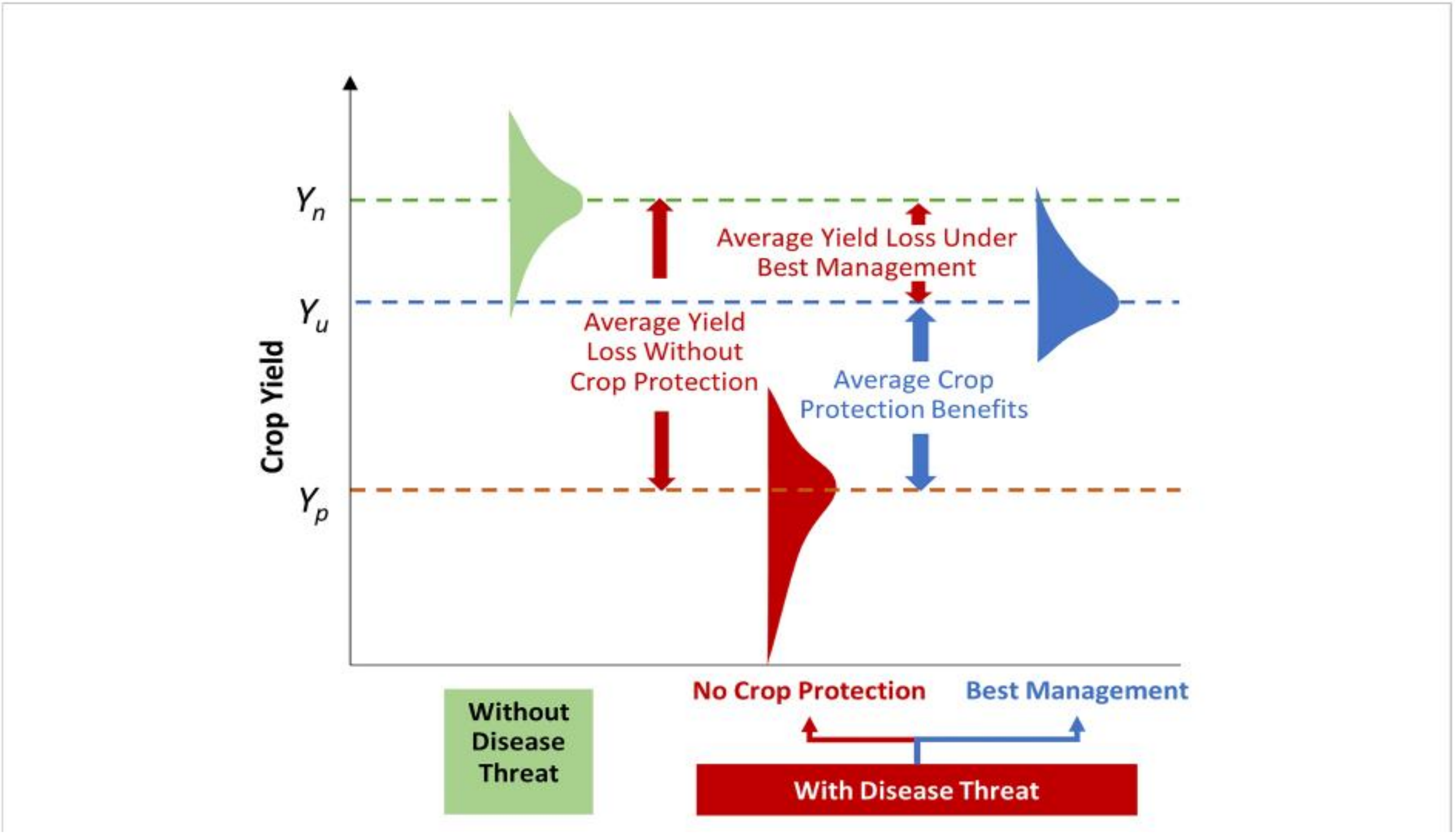
- Discussion questions (30 minutes)
- Review of conventional risk assessment tools (30 minutes)
  - Risk experiments vs revealed preference models
  - Mean-variance analyses
  - Stochastic dominance
  - Risk preference concepts: expected utility vs prospect theory, parametrized outcome space vs state contingent
- Advanced Tools
  - Spatial probabilistic assessment tool (1 hour)
    - Example 1:
      - ***Targeting:*** *Causal machine learning for precision CSA recommender system*
      - ***Probabilistic:*** *Spatial Bayesian methods primer*
  - Spatial risk assessment tool (1 hour)
    - Example 2:
      - *Spatial second order stochastic dominance and* ***risk aversion***
- Incorporating behavioural economics in CSA advisories (30 minutes)
  - Some suggestions

Discussion questions

Identify one CSA technology being advocated in area of interest

1. How do we incorporate risk and uncertainty in the current CSA advisories?
2. Does consideration of farmer risk preferences (some being risk lovers, risk neutral or risk averse) change what we would recommend?
3. How do we communicate uncertainty of our CSA advisories? Should we?
4. What risk mitigation (reduce losses) and risk protection (reduce shock occurrence) strategies should be bundled with the advisories?
5. Who are the winners and losers of the CSA advisories?

What do we mean by risk? Examples



Chai et al 2022

Review of risk assessment tools

- Direct survey questions
- Field experiments using multiple price list and related approaches
- Analysis of historical data using copula
- Mathematical farm optimization models with risk [static or dynamic stochastic programming]
  - Maximize system profits, minimize system risk
  - subject to: time constraints, labor constraints, land constraints
- Agent Based Models (ABM) with risk preferences and shocks among micro-level processes driving dynamic decision making.
- Approximations: Probabilistic (Monte Carlo Simulations) and risk assessment methods (Second Order Stochastic Dominance)

Some theoretical and empirical risk predictions

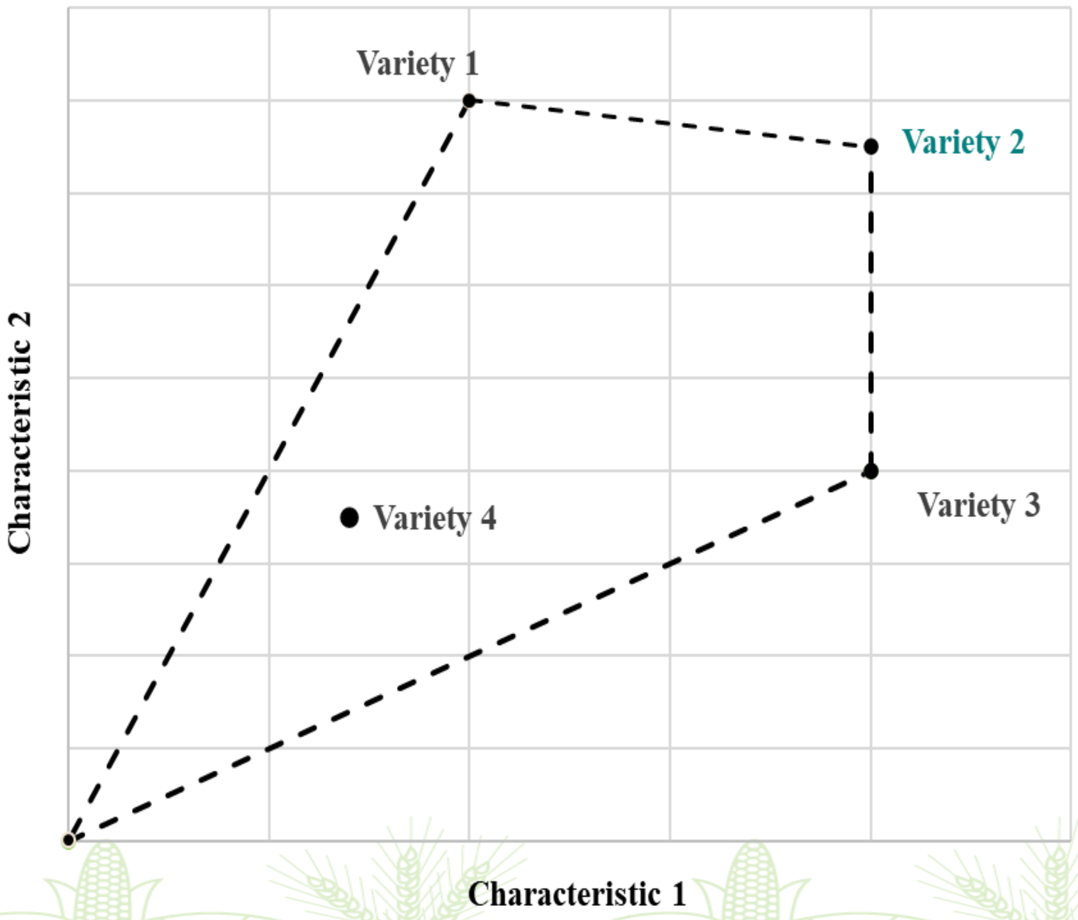
- Farmers response to output price risk
  - Sandmo (1971) and others after: Price risk at the extensive margin causes risk averse producers to decrease how much they produce. You can think of input recommendation implications!
  - Unfortunately no agreed theory on whether farmers increase particular inputs [Batra and Ullah, Hartman 1975]
- Consistently across many countries, there is low adoption of crop insurance
- Complementarities and substitutions in stochastic inputs may matter
- Some advisories have risk mitigation properties (e.g., irrigation, drought tolerant varieties) as such they are likely to be adopted in risk prone environments.

Portfolio frontier using mean-variance analysis

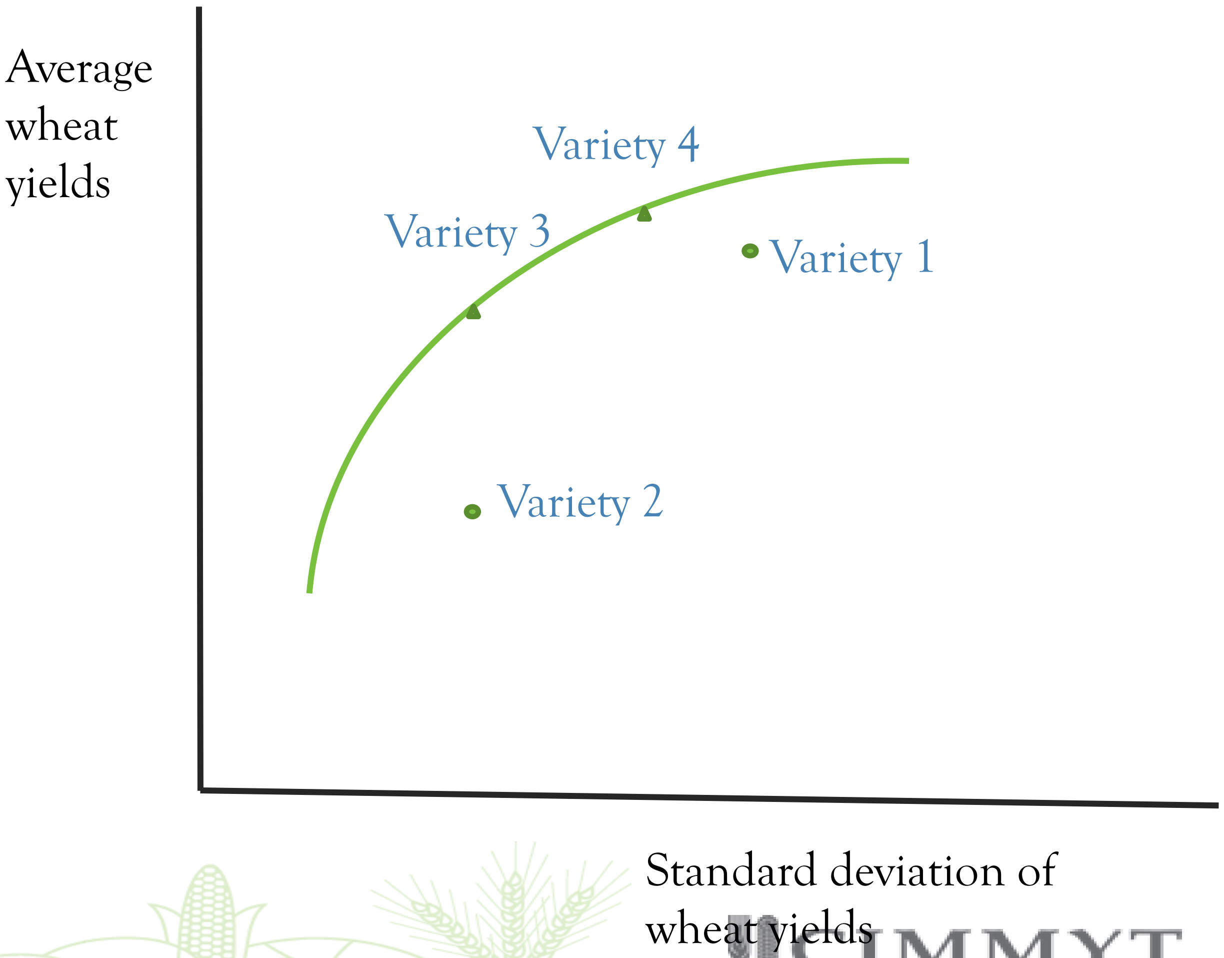
E-V analysis/Markowitz optimization

Varietal Portfolio frontier

Characteristics frontier



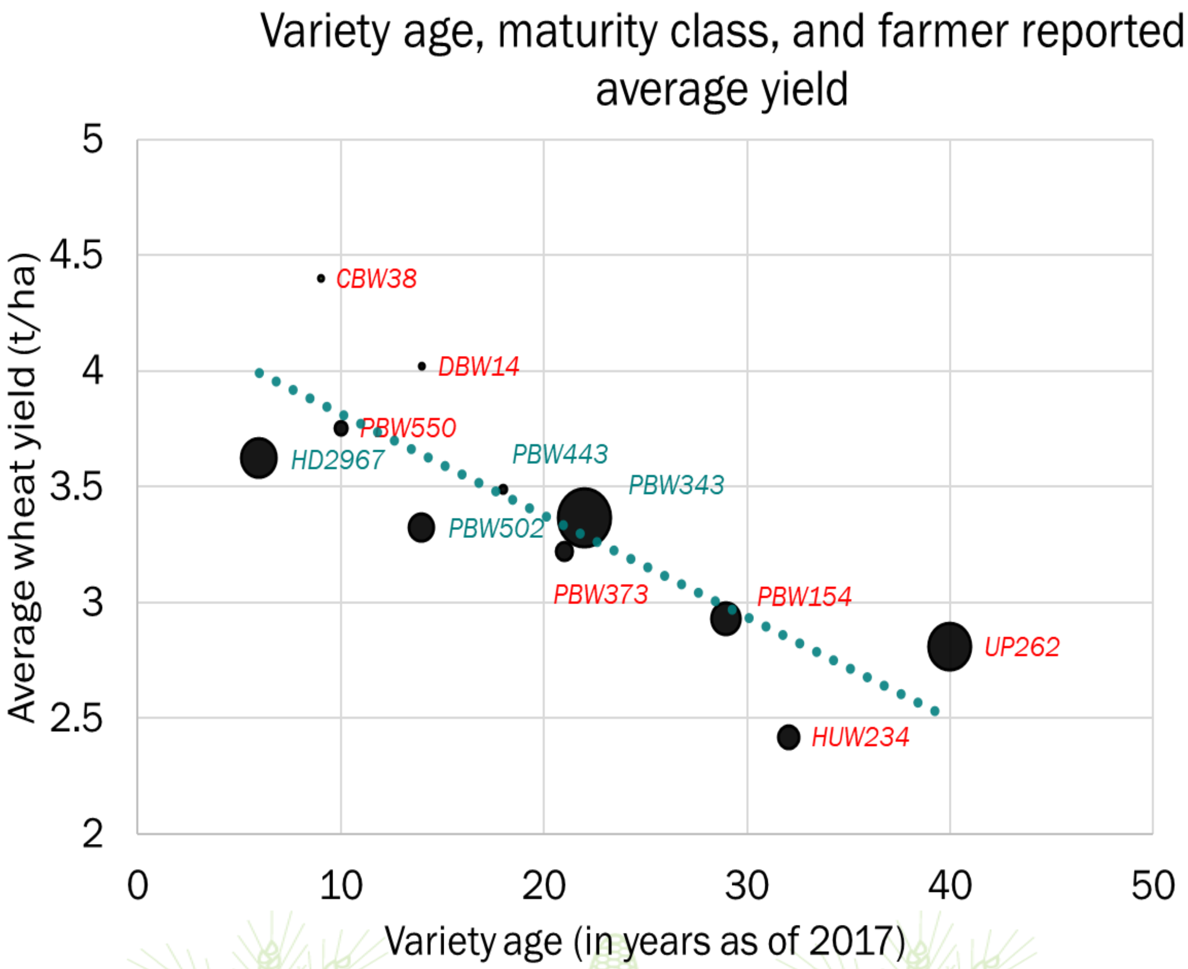
Portfolio frontier



How to find the frontier

- Visualization
- Pair wise comparisons and approximation
- Quadratic programming model (Excel Solver, quadprog and MNOF R packages)

Example (1): Survey Characteristics frontier



Legend

- Early
- Long duration (timely sown varieties)
- Size: adoption share

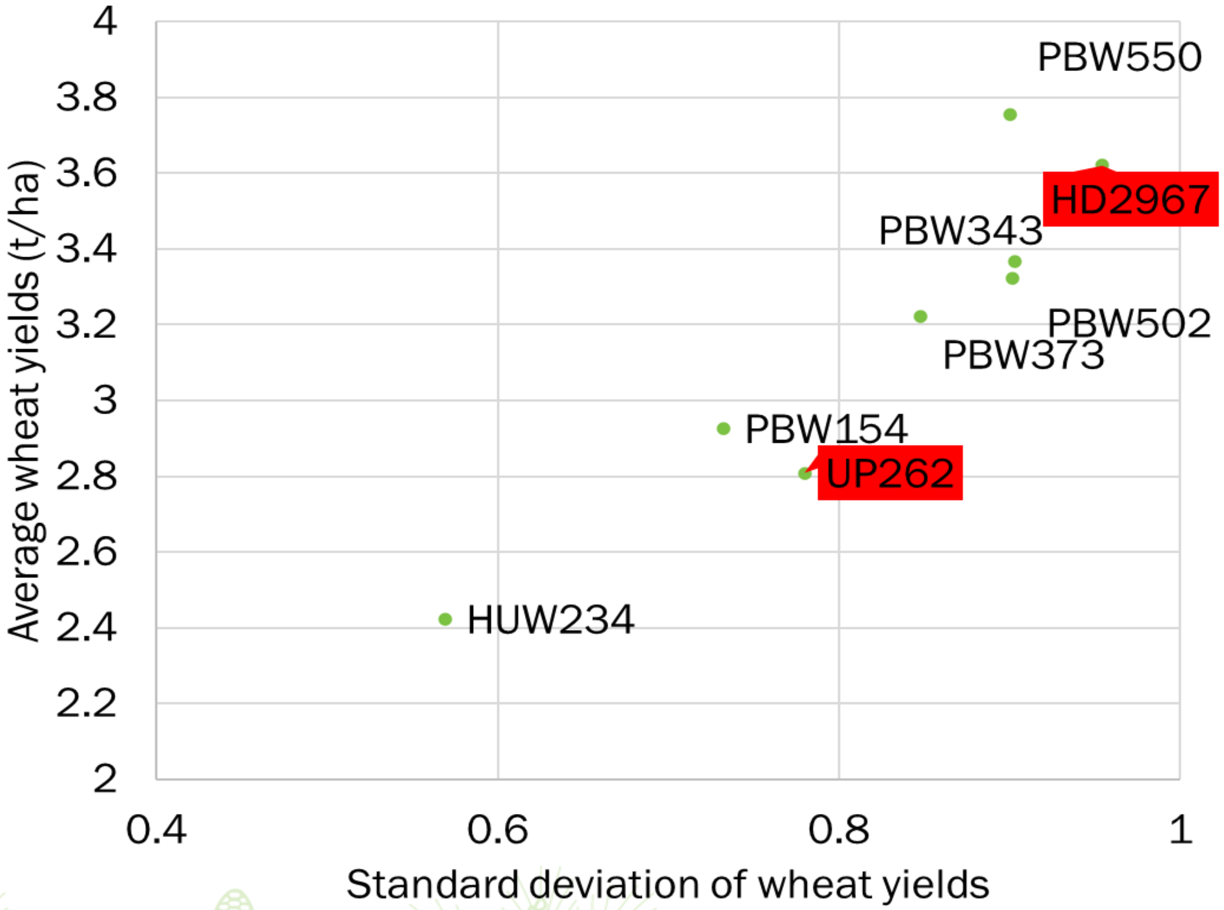
The Figure can be misleading

Why?

- Small sample sizes for less adopted varieties, e.g., CBW 38, DBW 14
- Conditions grown are different
- Some confounding traits like duration to maturity may matter more.
- We need to control for these

Example (2): Portfolio frontier

- By pairwise comparison, it makes sense to adopt PBW 154 as compared to UP 262
- More comprehensive comparisons: Mean-variance (E-V) optimization model using quadratic programming



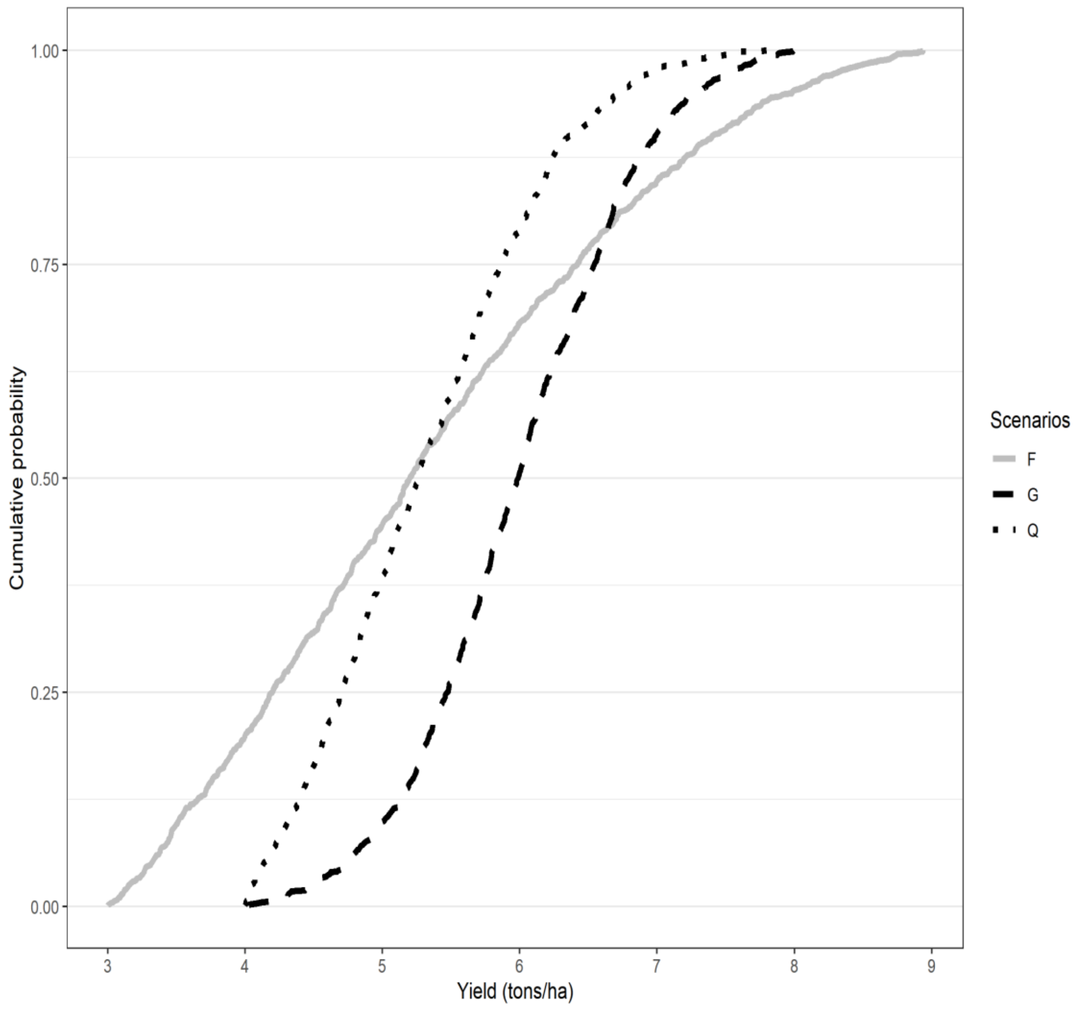
Stochastic Dominance Comparisons

Kolmogorov-Smirnov test

First and second order stochastic dominance

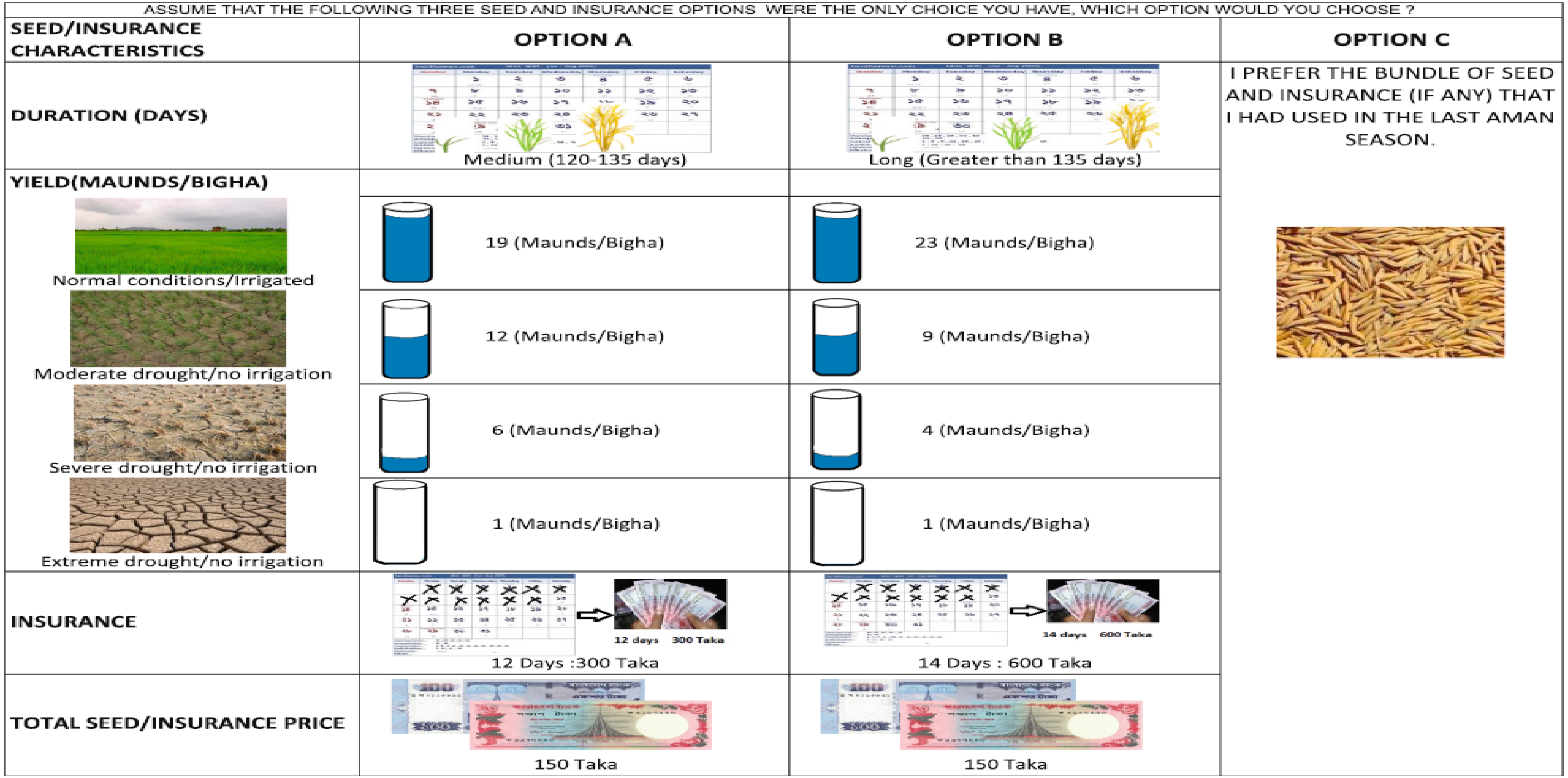
- First order stochastic dominance
  - G first order stochastically dominates Q
  - FOSD implies SOSD
- Second order stochastic dominance
  - Assuming same mean, agents prefer the technology with less outcome variance
  - Based on area of CDFs
  - G SOSD F

- G= rtruncnorm,
  - n=1000
  - a=4
  - b=8,
  - mean=6,
  - sd=0.8),
- Q=rtruncnorm
  - n=1000
  - a=4
  - b=8
  - mean=5
  - sd=1)
- F=rtruncnorm
  - n=1000
  - a=3
  - b=9
  - mean=5
  - sd=2.



CIMMYT

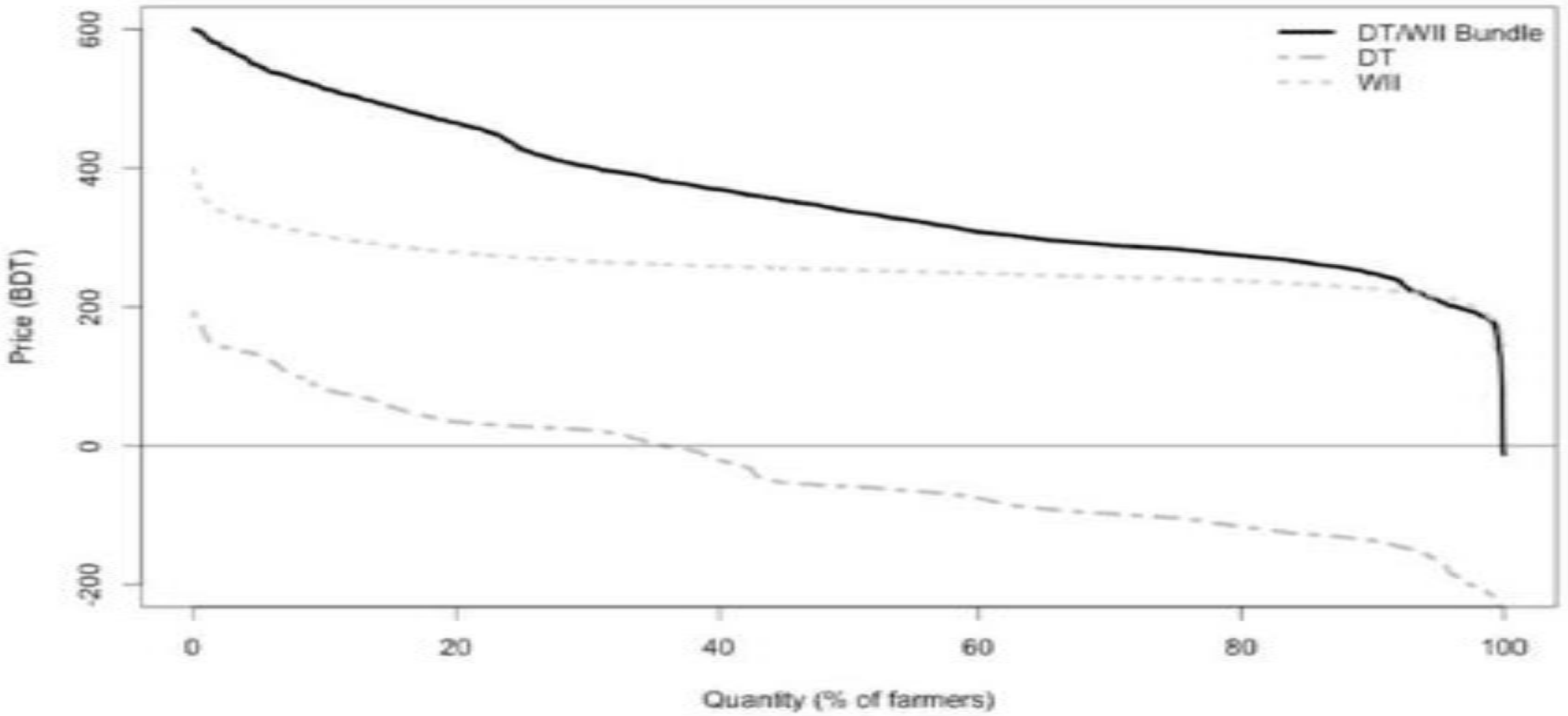
Example: Ward et al 2018
Demand for Complementary Financial and Technological Tools for Managing Drought Risk-Bangladeshi



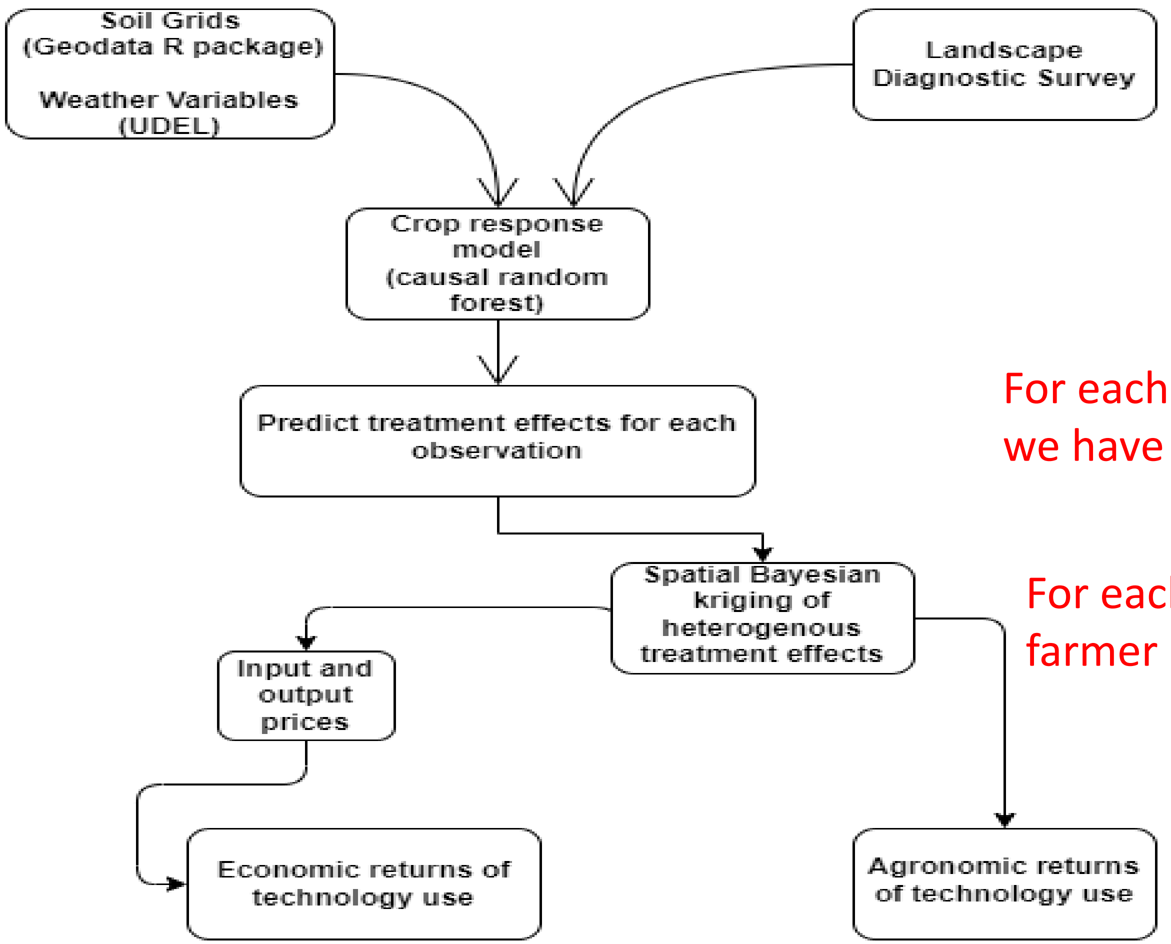
ASSUME THAT THE FOLLOWING THREE SEED AND INSURANCE OPTIONS WERE THE ONLY CHOICE YOU HAVE, WHICH OPTION WOULD YOU CHOOSE ?			
SEED/INSURANCE CHARACTERISTICS	OPTION A	OPTION B	OPTION C
DURATION (DAYS)	Medium (120-135 days)	Long (Greater than 135 days)	I PREFER THE BUNDLE OF SEED AND INSURANCE (IF ANY) THAT I HAD USED IN THE LAST AMAN SEASON.
YIELD(MAUNDS/BIGHA)			
Normal conditions/Irrigated	19 (Maunds/Bigha)	23 (Maunds/Bigha)	
Moderate drought/no irrigation	12 (Maunds/Bigha)	9 (Maunds/Bigha)	
Severe drought/no irrigation	6 (Maunds/Bigha)	4 (Maunds/Bigha)	
Extreme drought/no irrigation	1 (Maunds/Bigha)	1 (Maunds/Bigha)	
INSURANCE	12 days 300 Taka 12 Days :300 Taka	14 days 600 Taka 14 Days : 600 Taka	
TOTAL SEED/INSURANCE PRICE	150 Taka	150 Taka	

Fig. 2. Example of illustrated choice card presented to survey participants. Notes: 1 Maund is approximately equal to 40 kg. 1 Bigha is approximately equal to 1/3 acre. 1 Bangladeshi Taka (BDT) = 0.0129 USD at the time of the study. Choice cards shown to study participants were translated into the local language (Bangla).

Example: Willingness to pay for drought risk mitigation measures



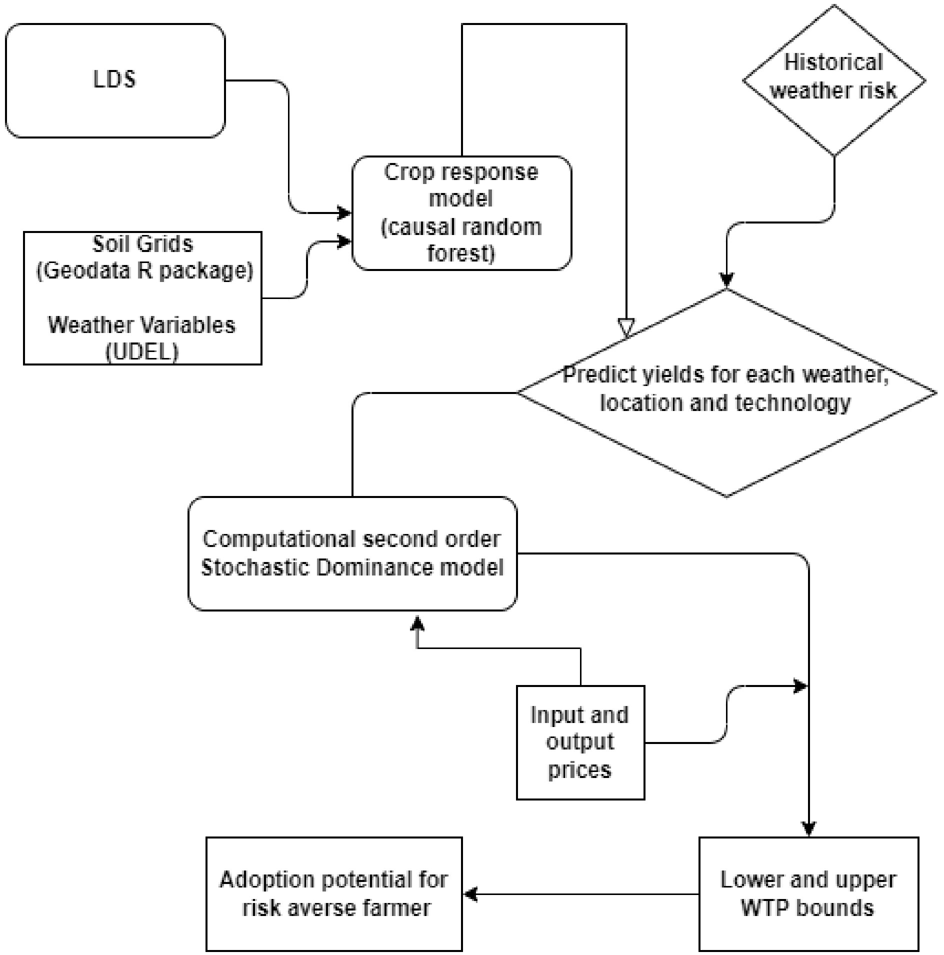
New Toolkits: Probabilistic and Risk Workflows



For each farmer we have data for

For each new farmer

Probabilistic assessment



Risk assessment

If it still beneficial to risk averse farmer, then recommend

Spatial Probabilistic Assessment

Spatially varying coefficient Bayesian linear model of causal random forest results

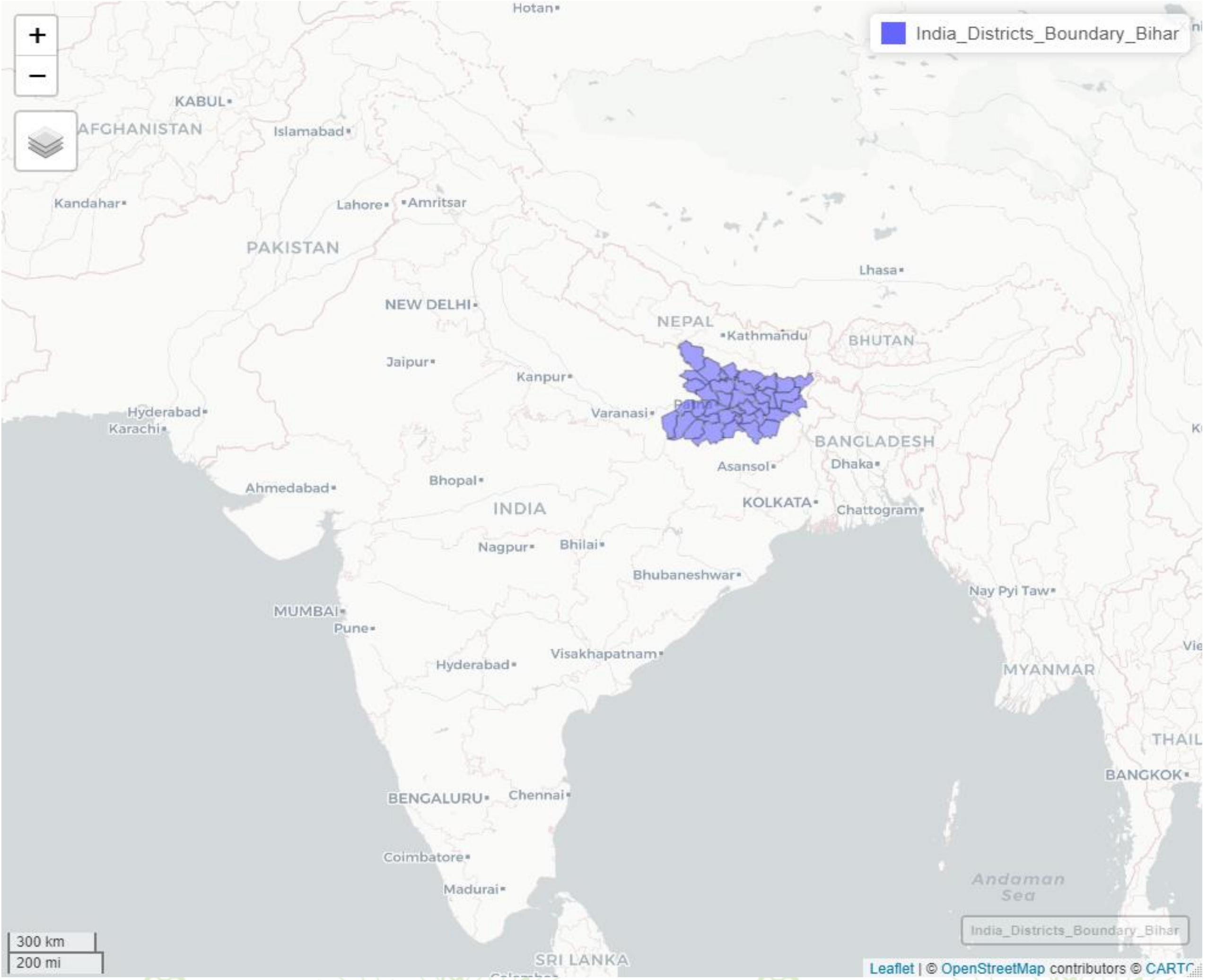
Best linear projection of treatment effects across space

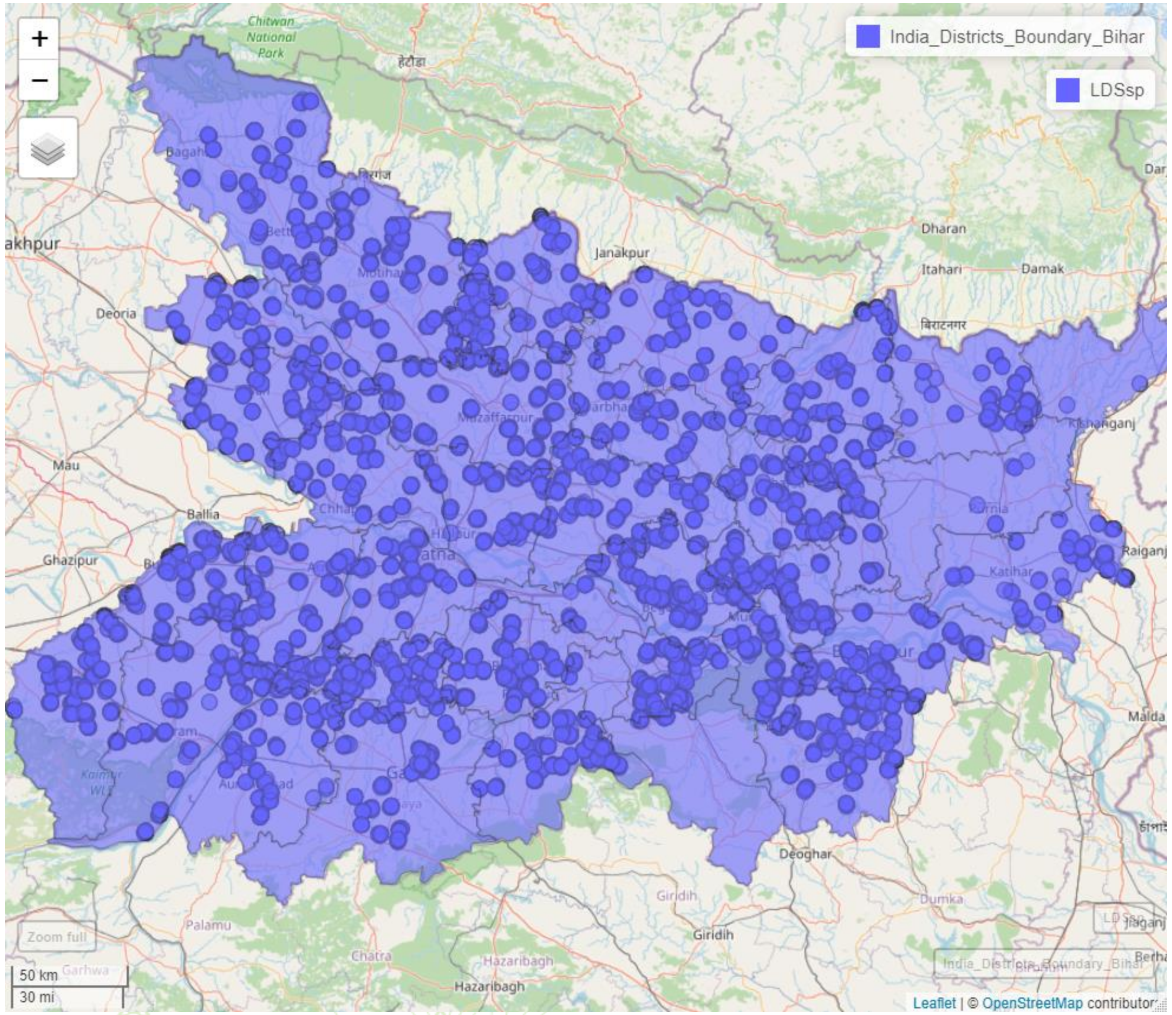
Data

- Data sources
  - Plot geocoded Landscape Diagnostic Survey (LDS) for wheat collected in 2017/18 [free online]:
    - ✓ This study: 8500+ irrigated wheat households in Bihar
  - Soil GRIDS
  - University of Delaware climate data (rainfall and temperature)
  - Government of India: Commission of Agricultural Costs and Prices
- Response variable: Wheat yield (t/ha)
- Genotypes: Variety type (short or long duration)
- Environment: Elevation, ruggedness, slope, precipitation, temperature, soil grids (soil pH, soil nitrogen, sand, e.t.c)
- Management: Sowing date (early or late), weed management, nutrient management (N, P, K applied), irrigation management (1 versus more irrigations) ,variety type (long versus short duration)
- Socio-economics: Distance to market, population density, gender, education, and prices.

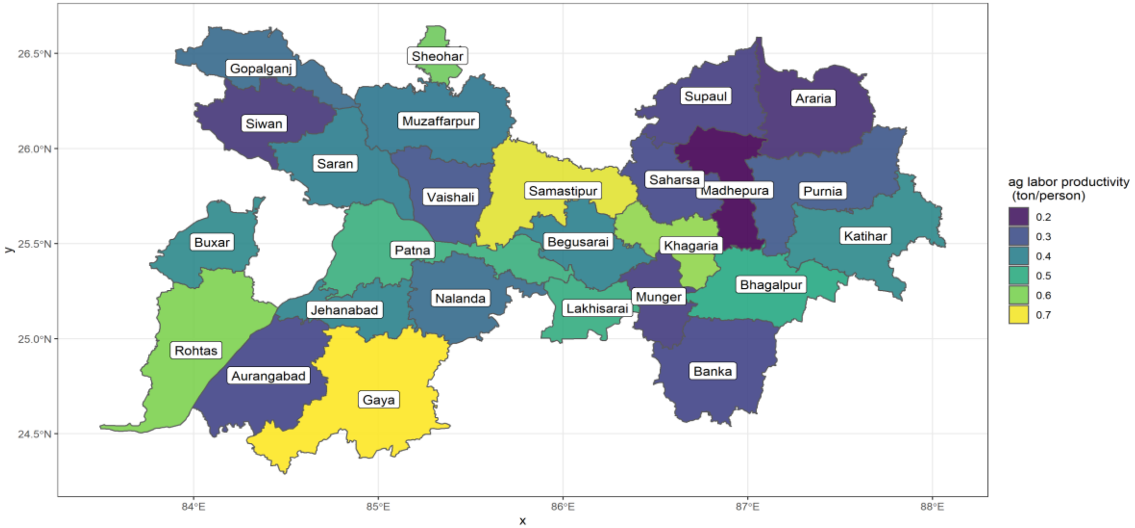
CIMMYT

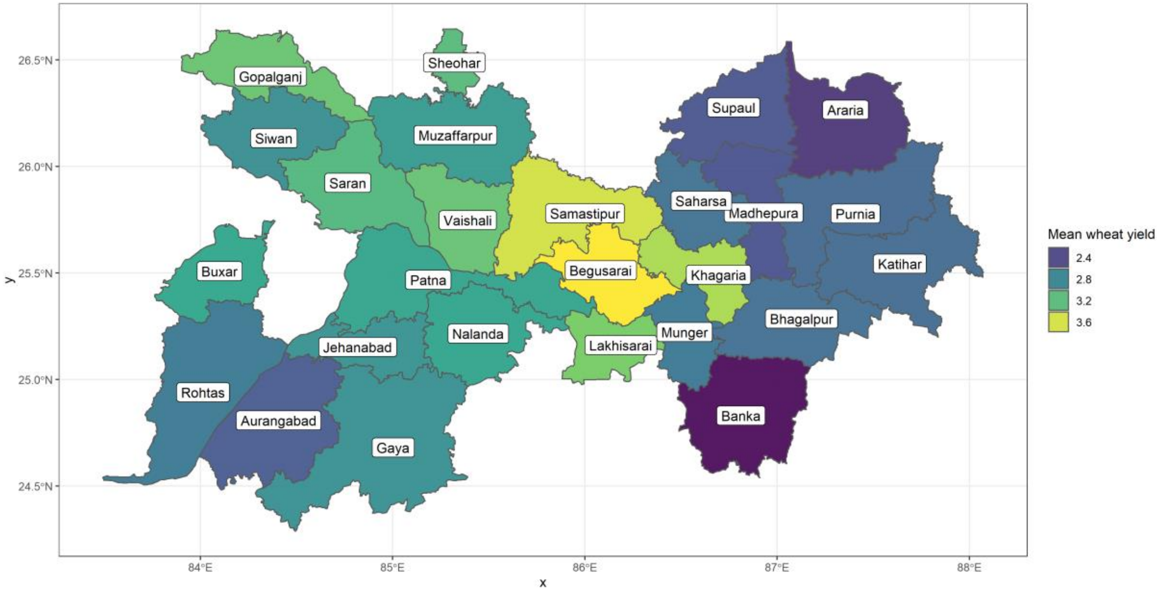
Bihar survey locations

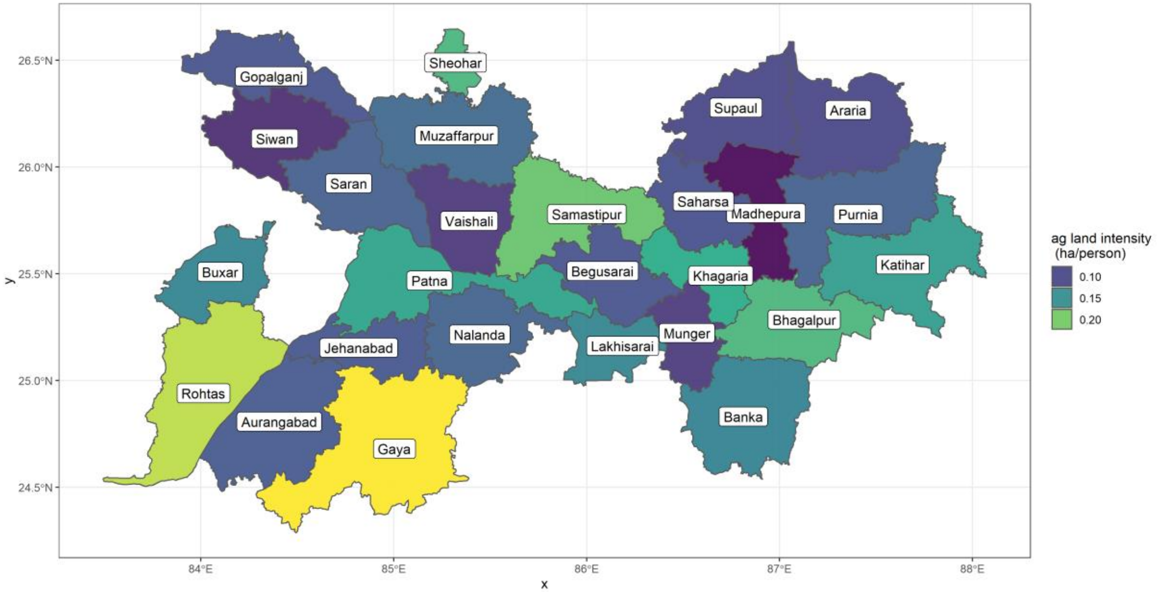




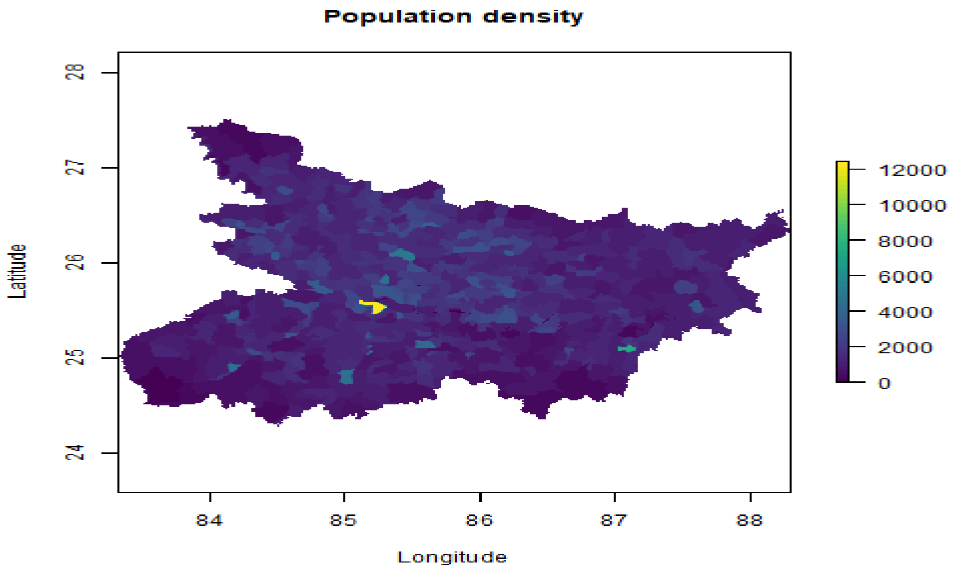
Wheat yield, labor productivity and land intensity by districts

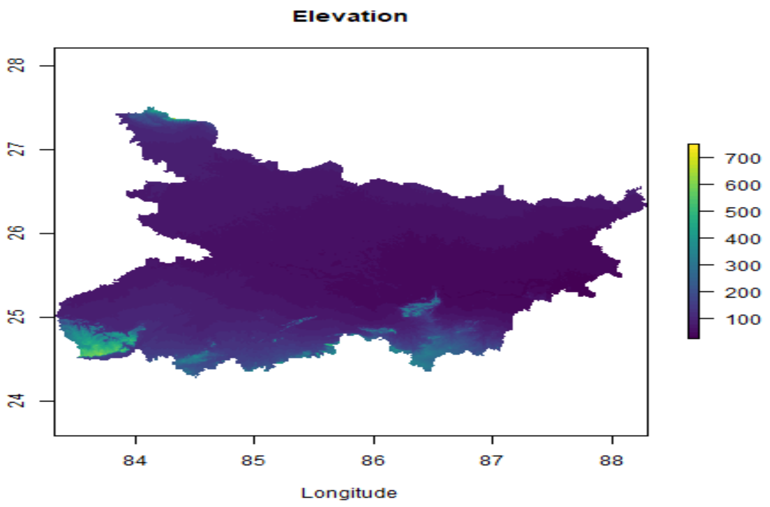


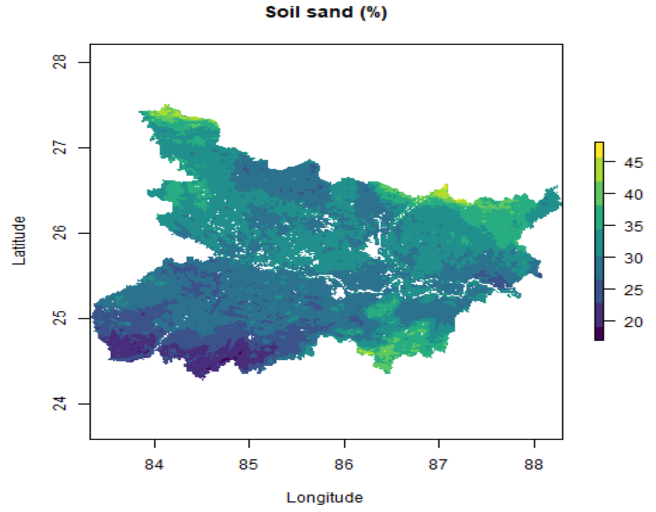


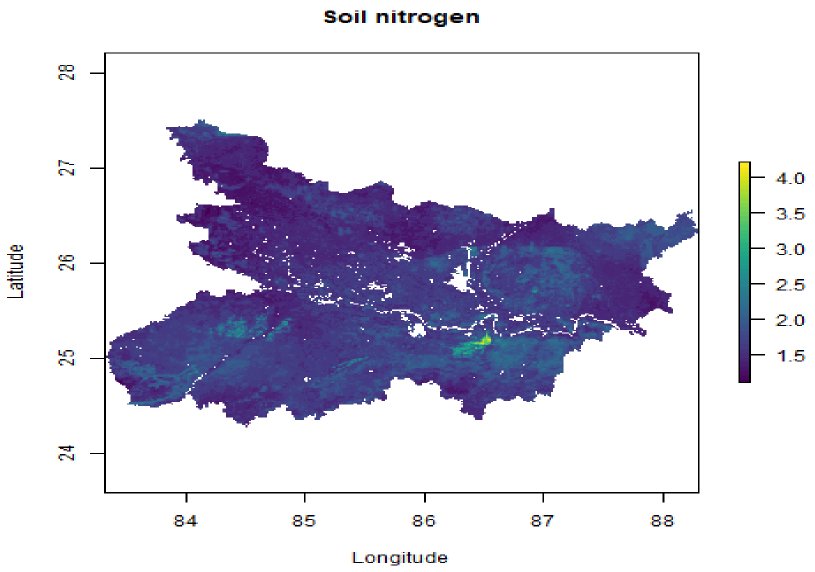


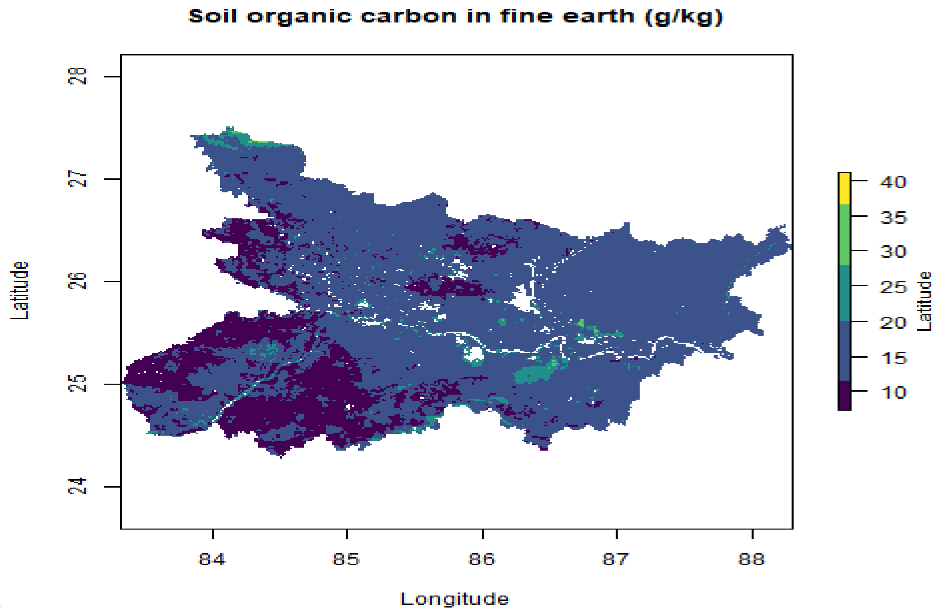
Some Covariates: from geodata R package

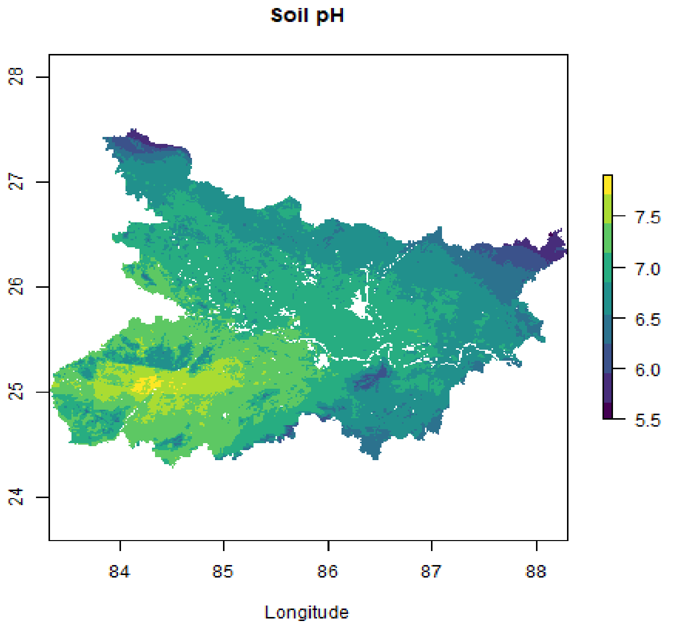












Empirical methods

- **Step 1:** Estimate a yield crop response model using causal random forest model with inverse weighted propensity score (AIWP)
- **Step 2:** Predict conditional average treatment effects (CATEs) for each location
- **Step 3:** Spatial Bayesian kriging of heterogenous treatment effects (can use Inverse Distance Weighted Interpolation if not concerned with uncertainty) .
  - Produce gridded maps showing yield responses, economic responses, and probabilities of getting gains above some threshold.
- **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.

Primer(1): from OLS to spatial +ML

- Data: (Y_i, X_i, s_i): $i = 1, \ldots n$
  - Y_i: wheat yield
  - X_i: Genotypes, Environment, Management, Socio-economic covariates
  - s_i: Location
- Areal location: Spatial error, spatial lag, spatial durbin model (to some this is all spatial econometrics, wrong!).
- Point location: Geostatistical/geospatial models
- Aim:
  - Understand relationship between X and Y [ideally causal relationship]
  - Predict at a new location s_0
- Ordinary Least Squares (OLS)

$$Y = \beta_0 + \beta_1 x + \epsilon$$

- Two problems: Assumes linear (in parameters) relationship, only spatially homogenous relationship can be assumed

CIMMYT

Primer(1): from OLS to spatial +ML

- Simple spatial linear mixed effect model

$$Y = \beta_0 + \beta_1 x(s) + \omega(s) + \epsilon$$

  - Problem: assumes linearity and homogeneity in β
- Simple spatially varying coefficient linear model

$$Y = \beta_0(s) + \beta_1(s)x + \omega(s) + \epsilon$$

  - Problem: assumes linearity
- Solution: Machine learning methods especially random forest
  - No need for **functional form debates** (on quadratic, cobb-Douglas, translog, linear-plateau, quadratic plateau, e.t.c.)
  - But what about **spatial correlation? Tobler's First Law of Geography?**
  - ✓ *"Everything is related to everything else, but near things are more related than distant things."*
  - What about **endogeneity problems** economists are always concerned with?

Why causal random forest instead of random forest?

- In random forests, the data is repeatedly split in order to minimize prediction error of an outcome variable.

- Causal forests (Athey et al 2019-Generalized Random Forests) are built similarly, except that instead of minimizing prediction error, data is split in order to maximize the difference across splits in the relationship between an outcome variable and a "treatment" variable.

- This is intended to uncover how treatment effects vary across a sample.

Why causal random forest?
(As shown by Kurmangaliyeva 2021)

- Some people wrongly use a Random Forest to simply learn two predictions and take the difference

$$\tau^{RF}(x) = \underbrace{\{\hat{y}^{(RF)}(W = 1, X = x) - \hat{y}^{(RF)}(W = 0, X = x)\}}_{\color{red}{Wrong}}$$

- Even under an RCT, $\hat{\tau}^{RF}(x)$ will be biased
- Why? Because we used the same data to decide how to partition into leaves and to estimate mean values within each leaf.
- Causal trees help resolve this problem using Honest splitting.

Primer: Causal random forest

- CRF as extension of heterogeneity analysis

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

- Potential outcomes framework

$$\widehat{\tau_i}(x) \equiv E(Y_i(1) - Y_i(0) | X_i = x)$$

- Honest splitting: Splitting using training sample and uses independent sample to estimate leaf means
- Orthogonalization for observational (survey) data using Inverse-Probability Weights predicted from probit model. AIPW-CATE is given by

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

Where $e(X_i)$ treatment assignment probabilities

This paper: Yield and economic returns

- Non-linear treatment estimation of yield returns [Causal Random Forest]

$$YR = \Delta y = \tau^{CRF}$$

- Economic returns

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

- Sensitivity of yield and economic returns
  - Best linear projection of the conditional average treatment effect

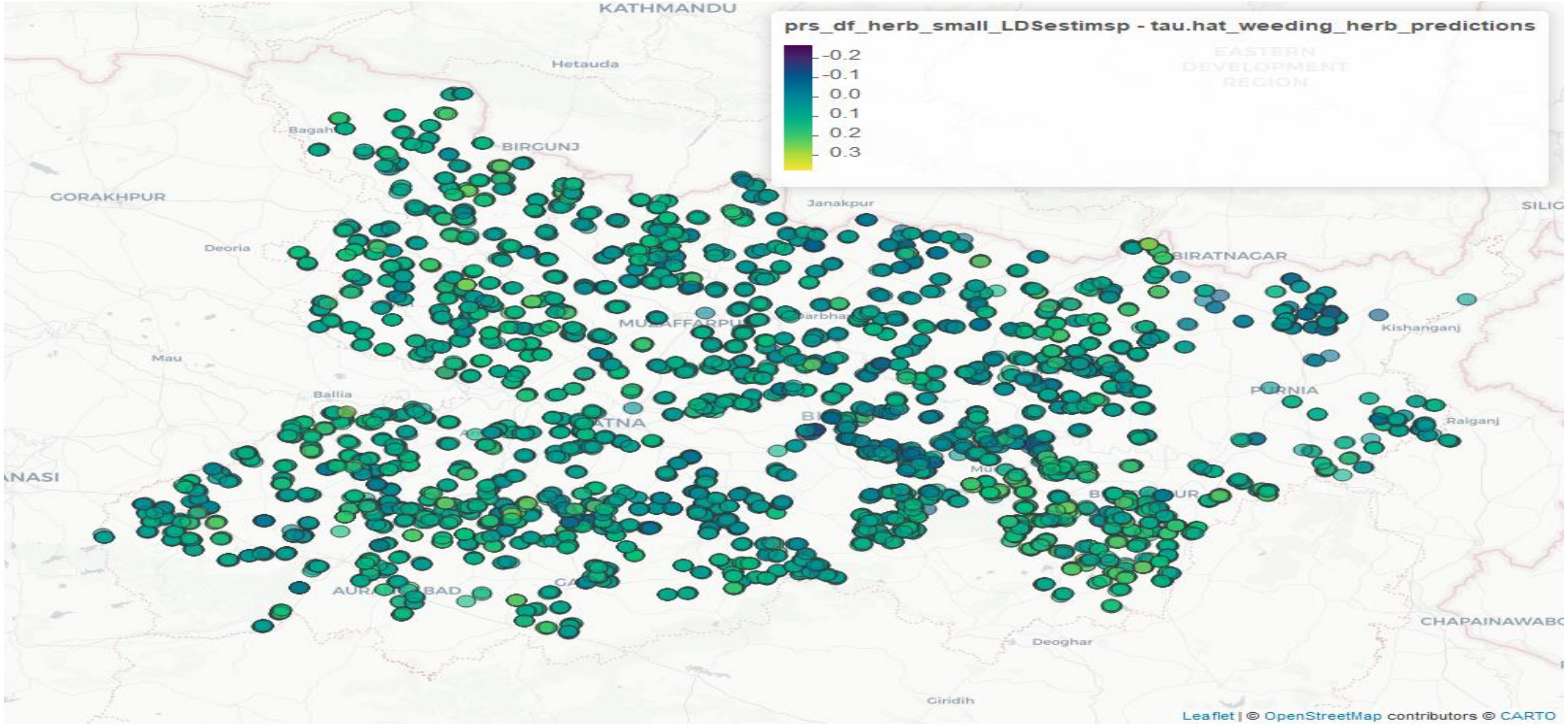
$$\tau^{CRF} = \beta_0 + \beta Z + \epsilon$$

- Spatial effect as second stage Gaussian process using Spatial Bayesian Linear Model

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

- where spatial effect is exponential GP, i.e., $C(s_i, s_j) = \sigma^2 \exp(-\phi \left\| s_i - s_j \right\|_2)$
- Predicting treatment effects to new locations with parameter uncertainty using spatial Bayesian kriging

Predicted yield gains (t/ha): Herbicide weed Management

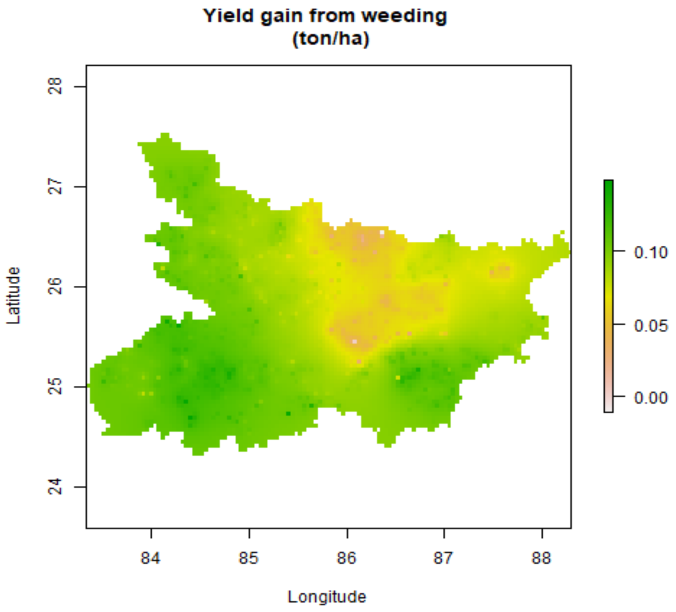


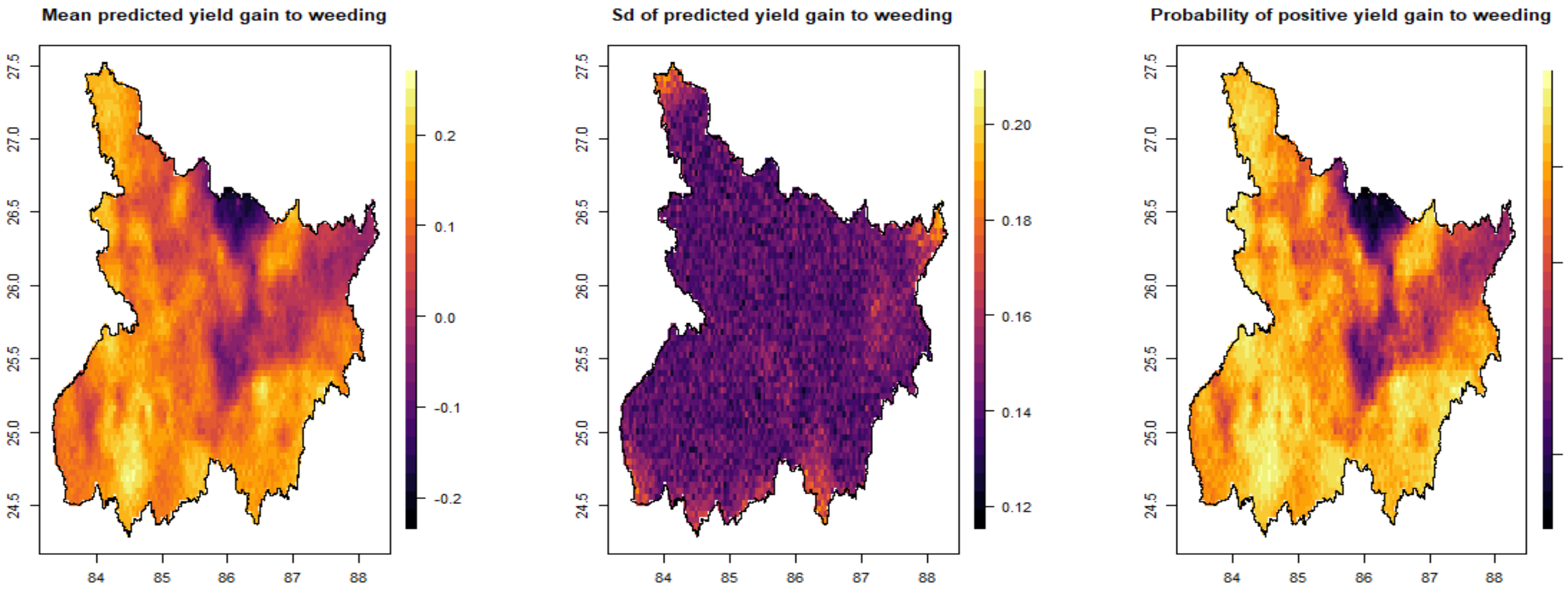
CIMMYT

Predicting into new locations: requires model that captures spatial correlation

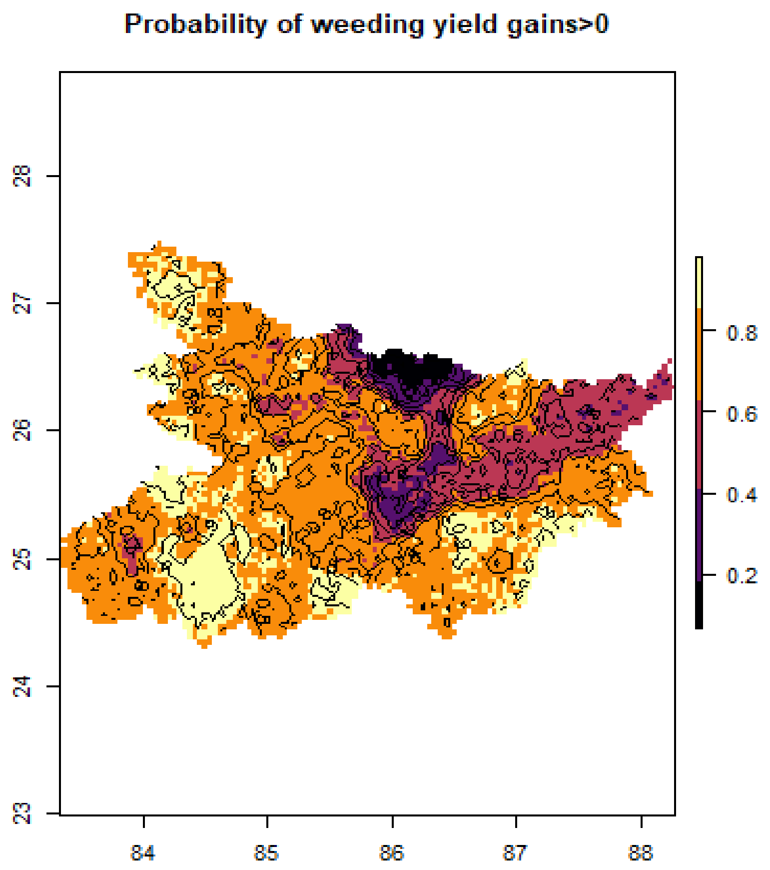
IDW interpolation CRF:
No statistical model assumed, so no variance estimate

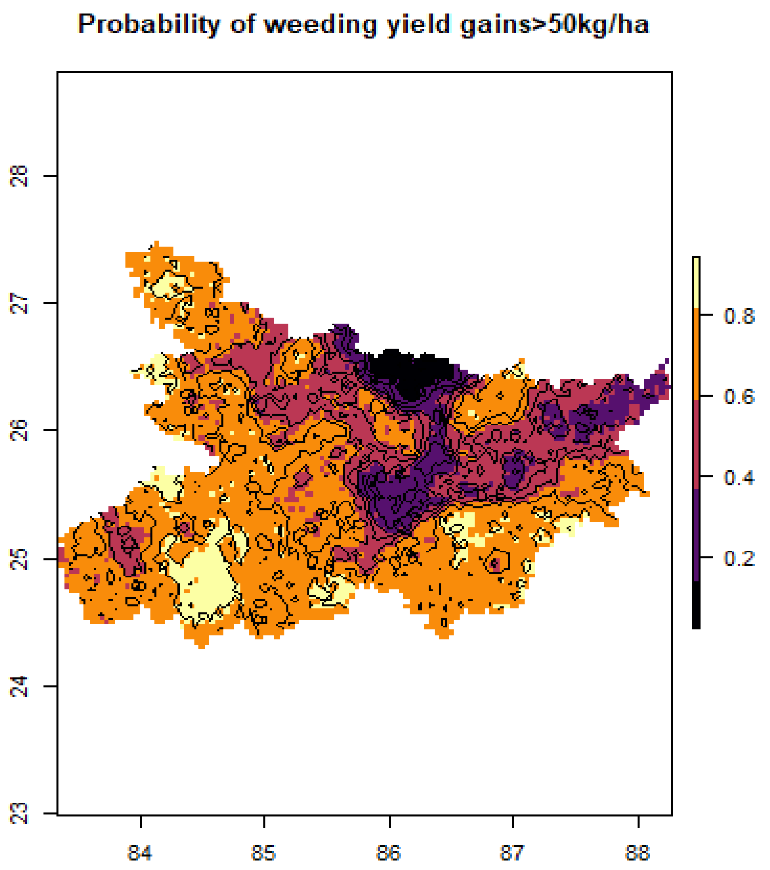
Spatial Bayesian kriging of CATEs: allows prediction uncertainty

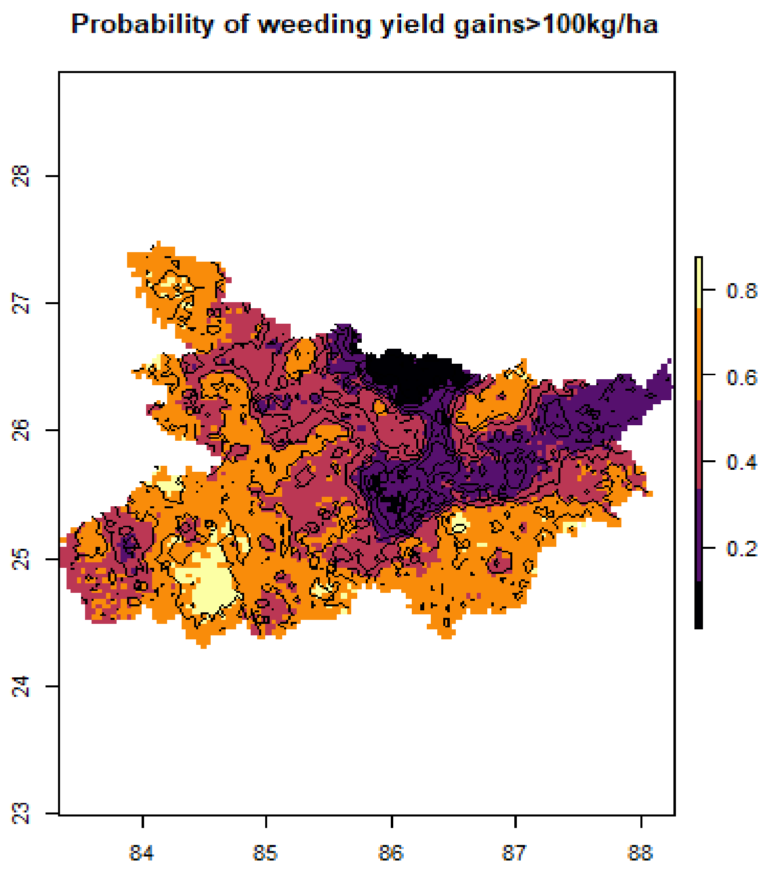




Probability of getting yield gains above a threshold

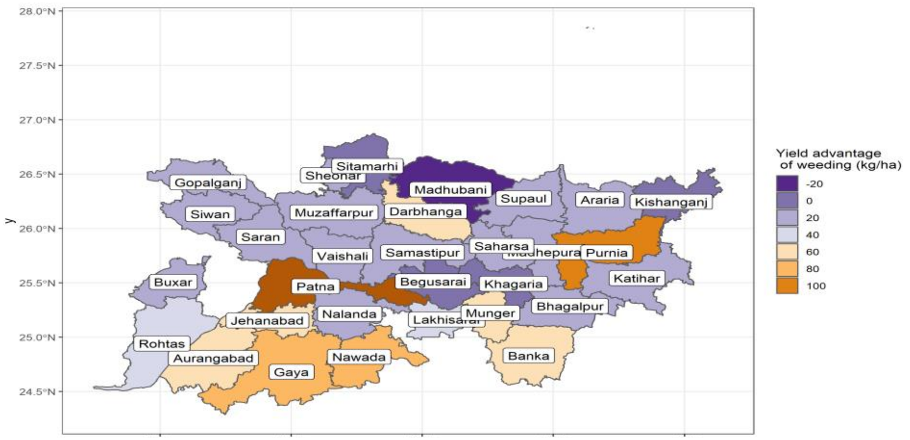


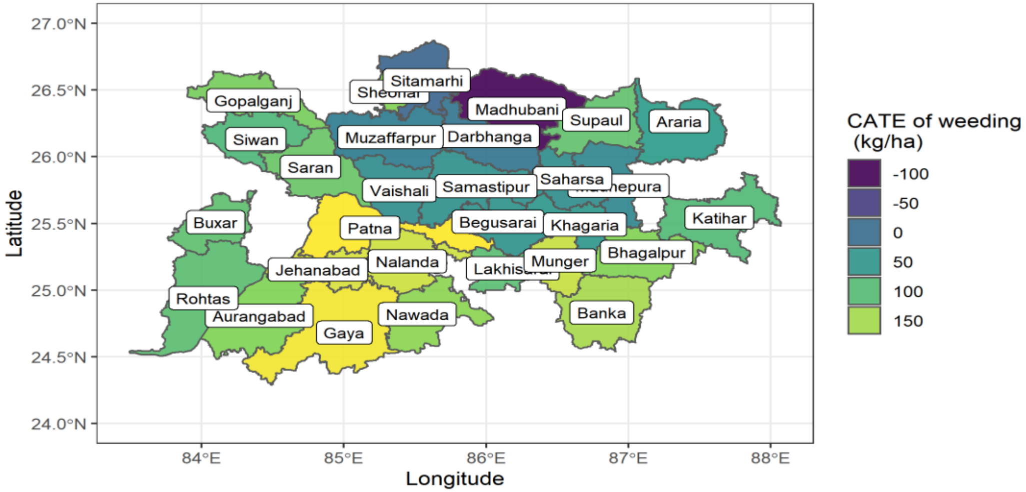


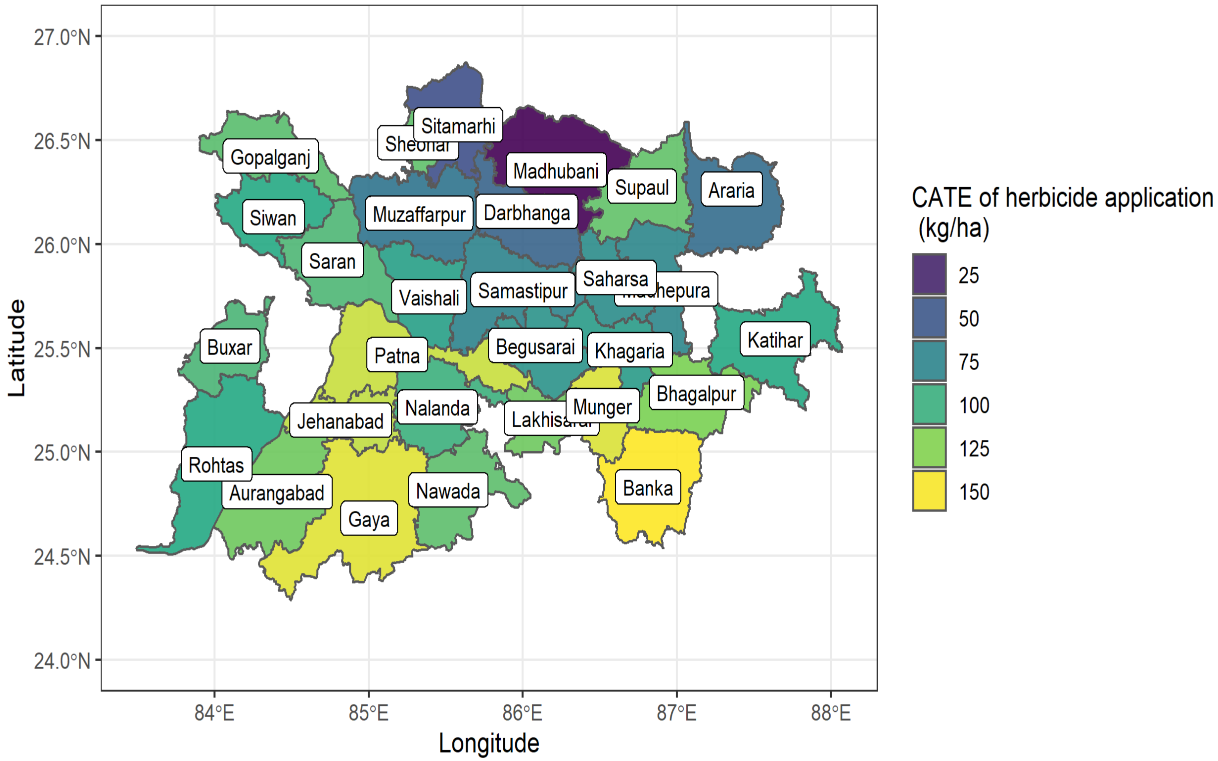


CIMMYT

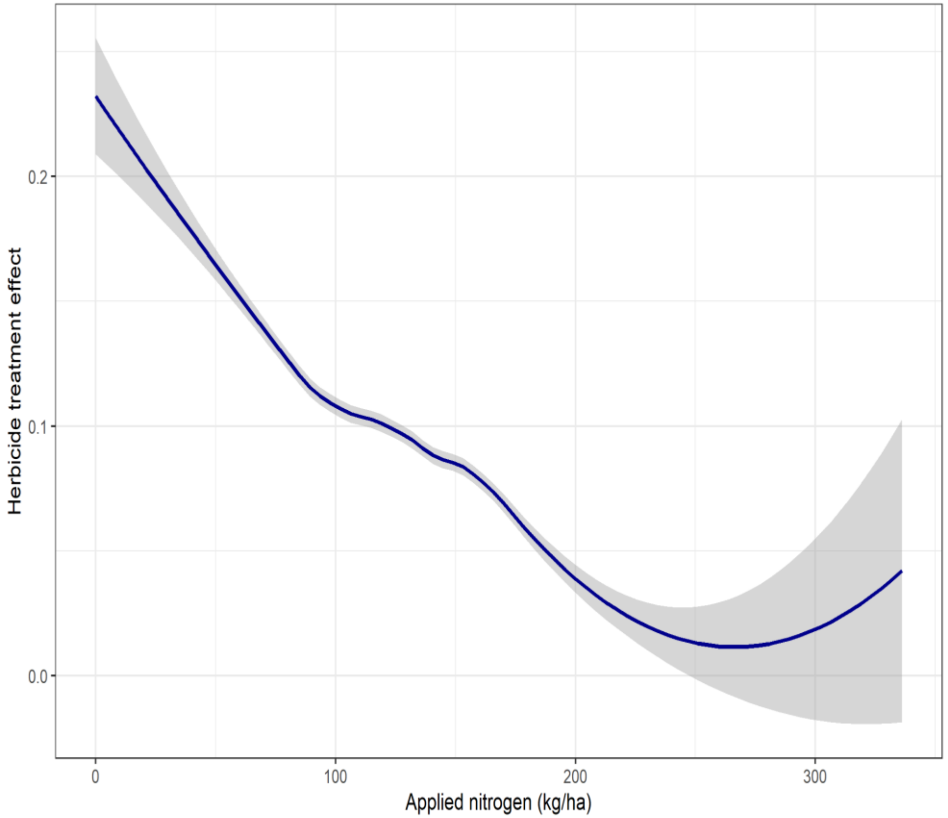
Heterogeneous Treatment Effects (HTEs) from causal random forest model

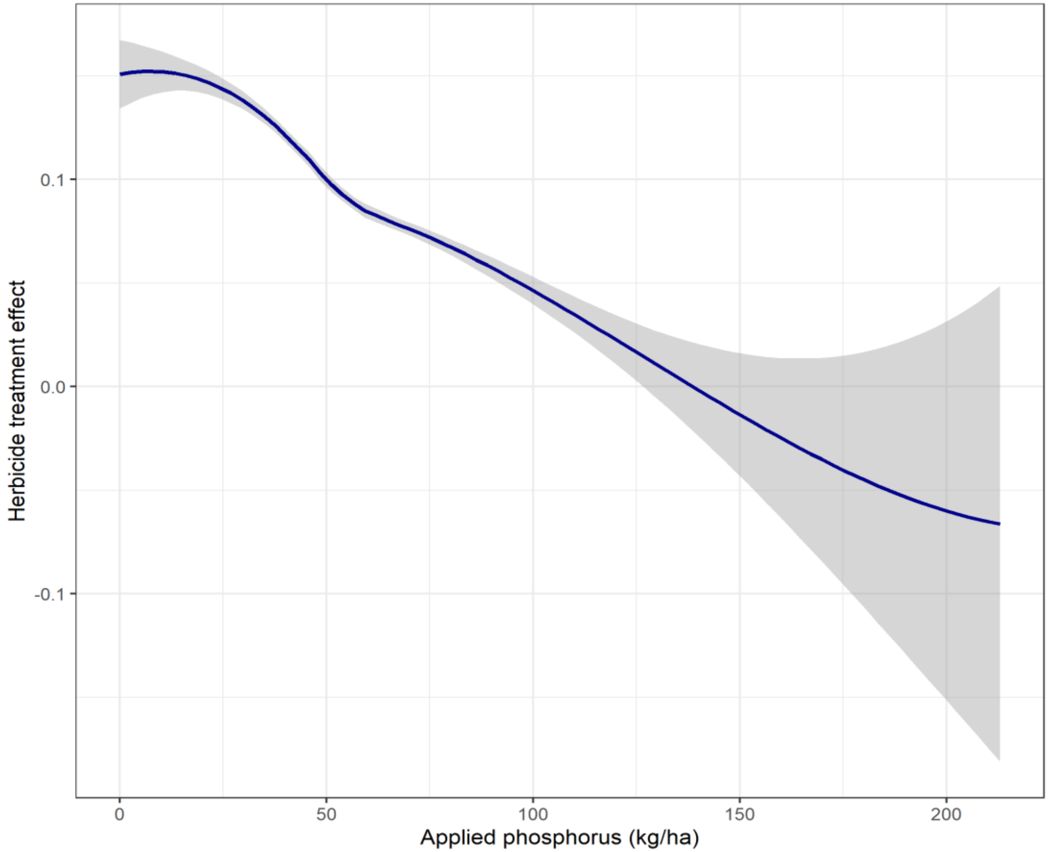






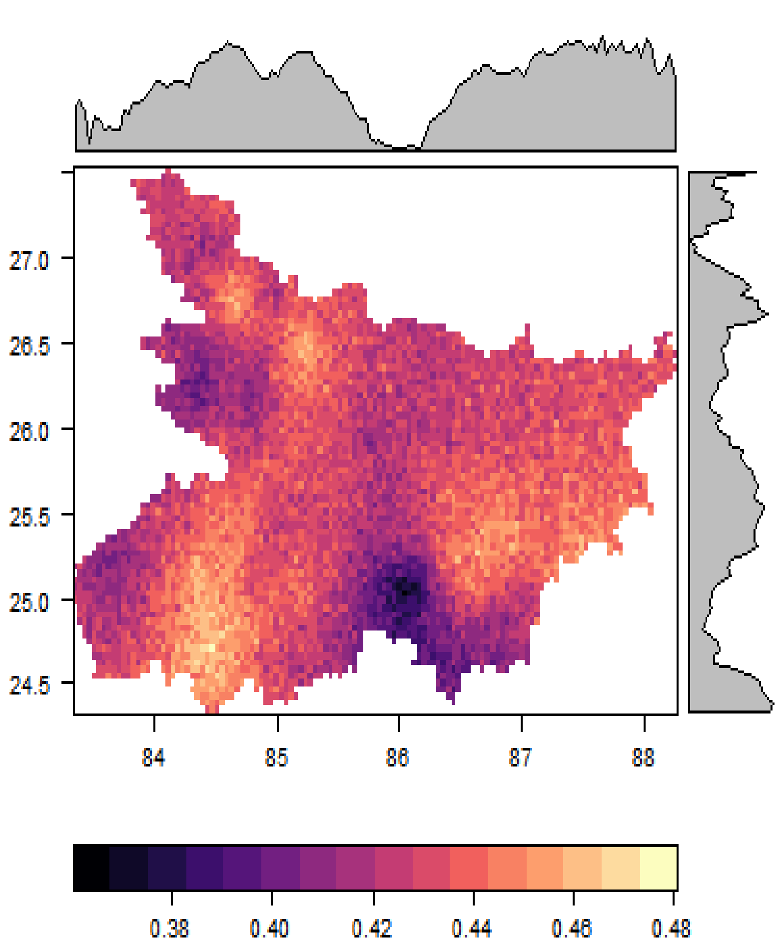
From CATEs to GATES
(Group Average Treatment Effects)



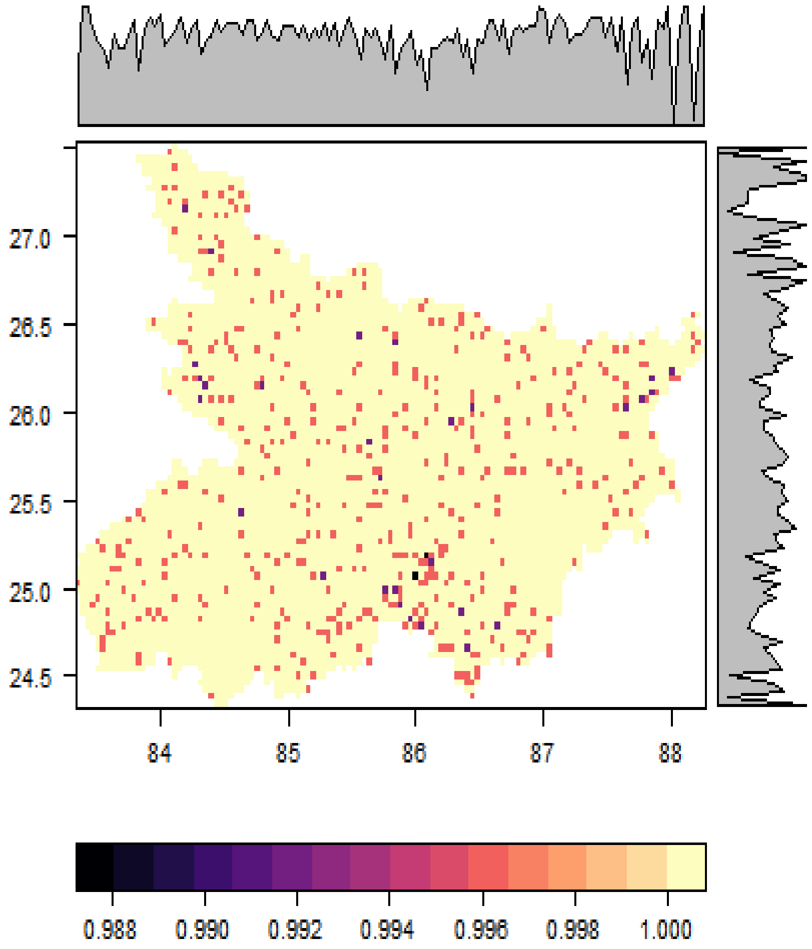


Other applications: Two vs. 1 irrigation yield gains (t/ha)

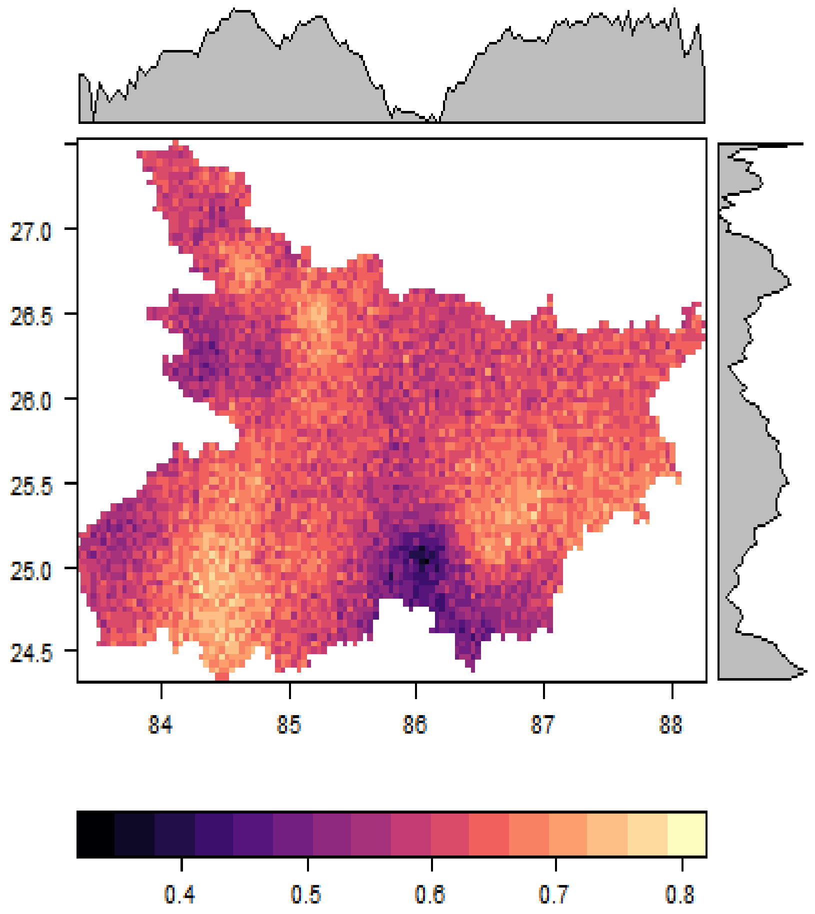
Average yield gains



Probability to get >100 kg/ha yield gain



Probability to get >400kg/ha yield gain

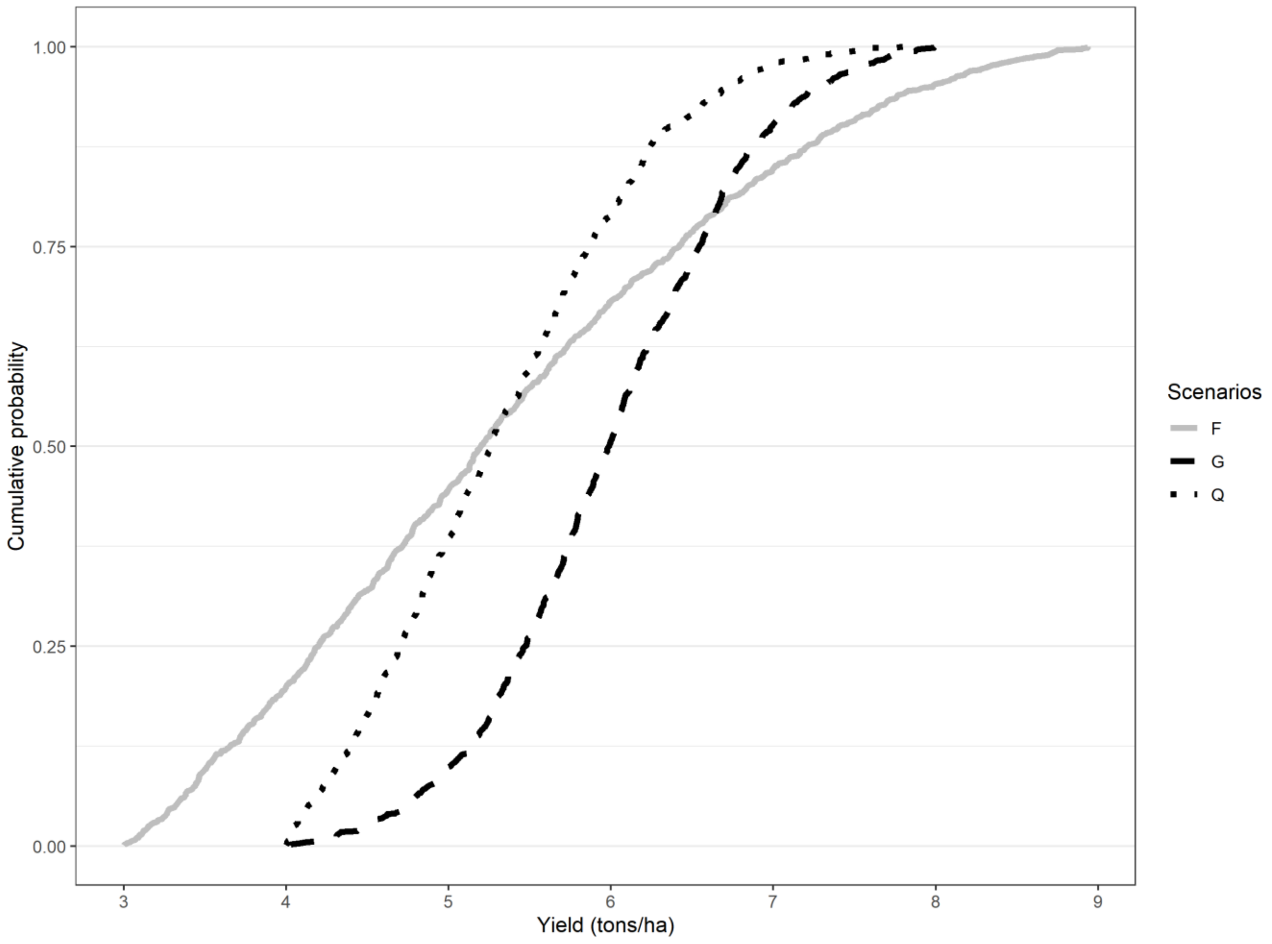


Spatial Risk Assessment

Using relationship between second order stochastic dominance and risk aversion

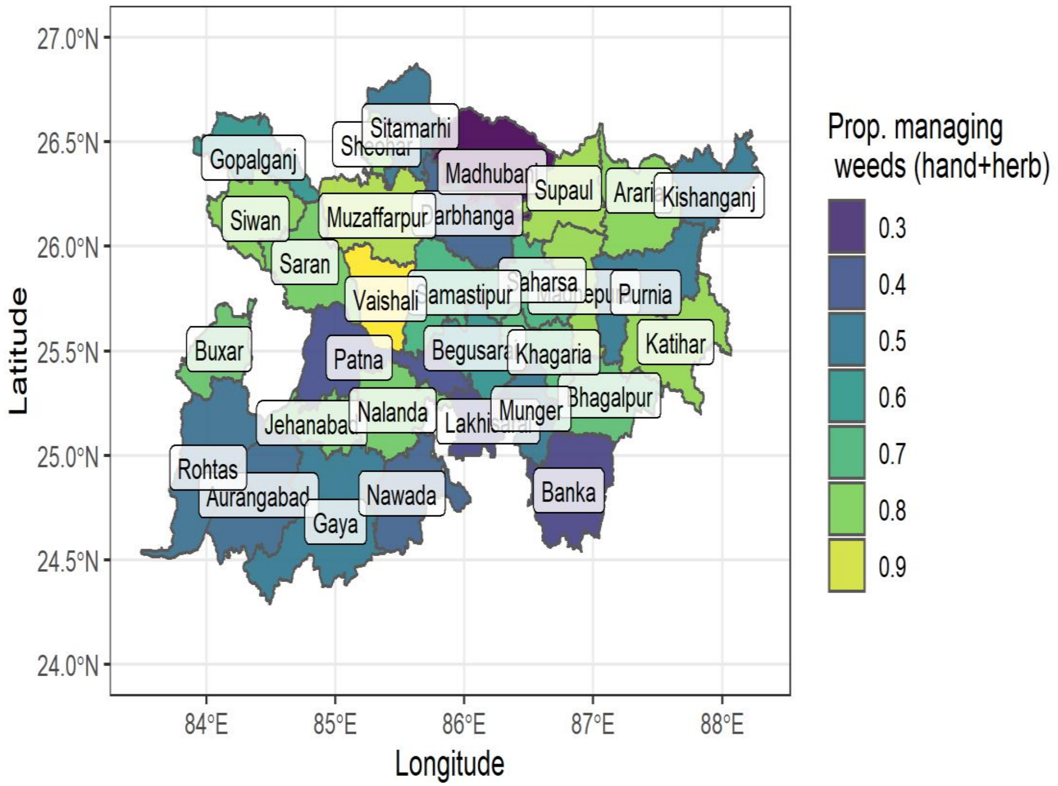
Key reference: Hurley et al 2018

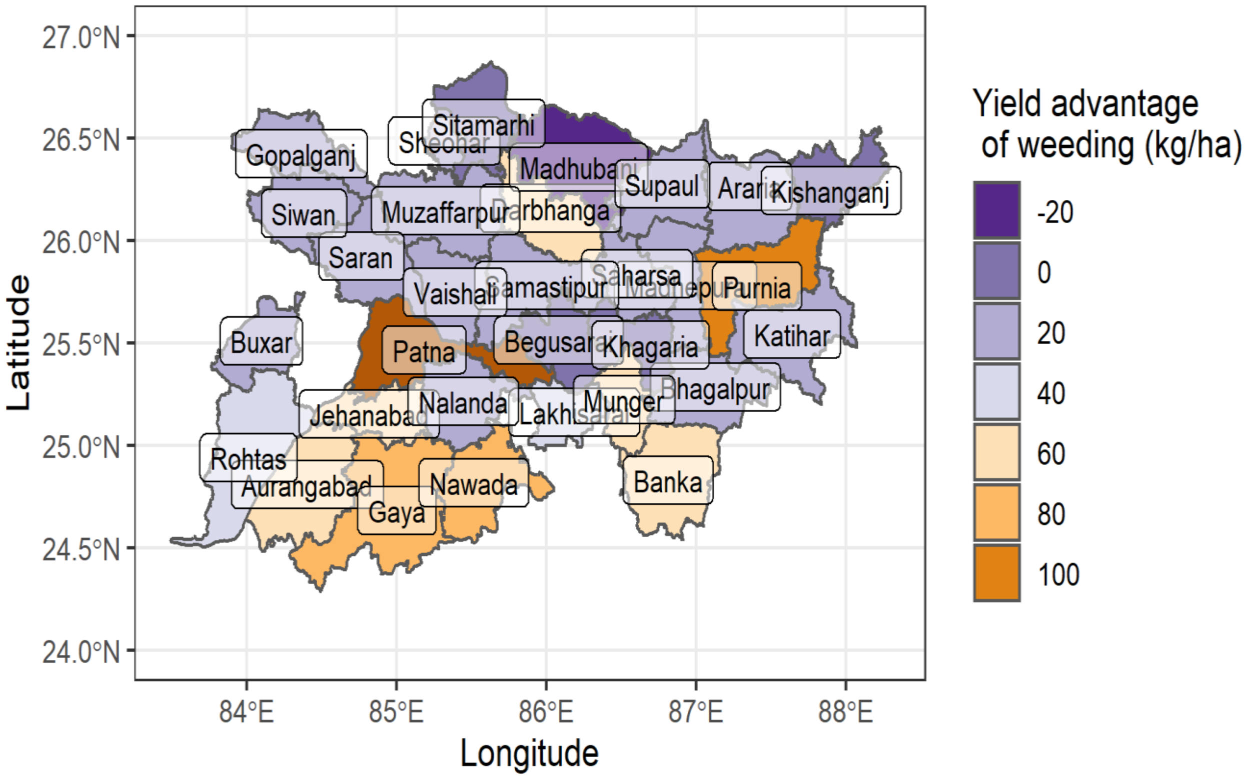
Risk assessment approach: Second order stochastic dominance



- G= rtruncnorm,
  - n=1000
  - a=4
  - b=8,
  - mean=6,
  - sd=0.8),
- Q=rtruncnorm
  - n=1000
  - a=4
  - b=8
  - mean=5
  - sd=1)
- F=rtruncnorm
  - n=1000
  - a=3
  - b=9
  - mean=5
  - sd=2.

Adoption, and random forest yield advantage of weeding





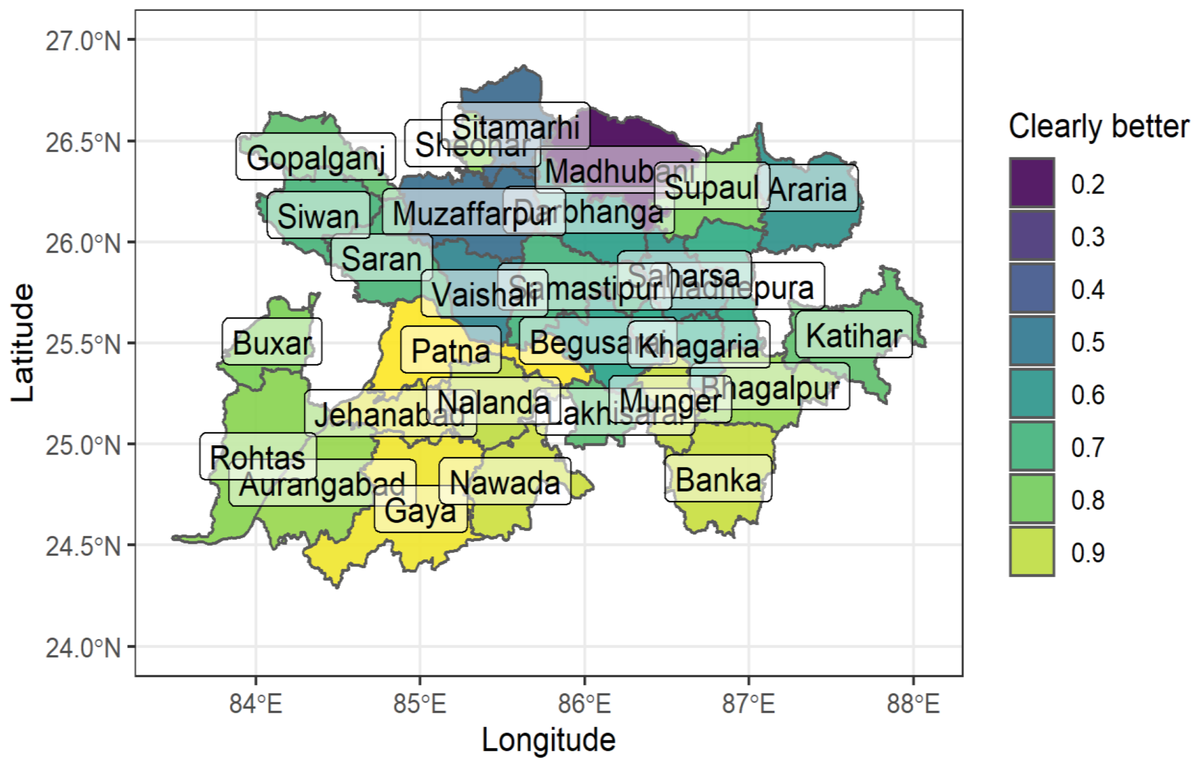
CIMMYT

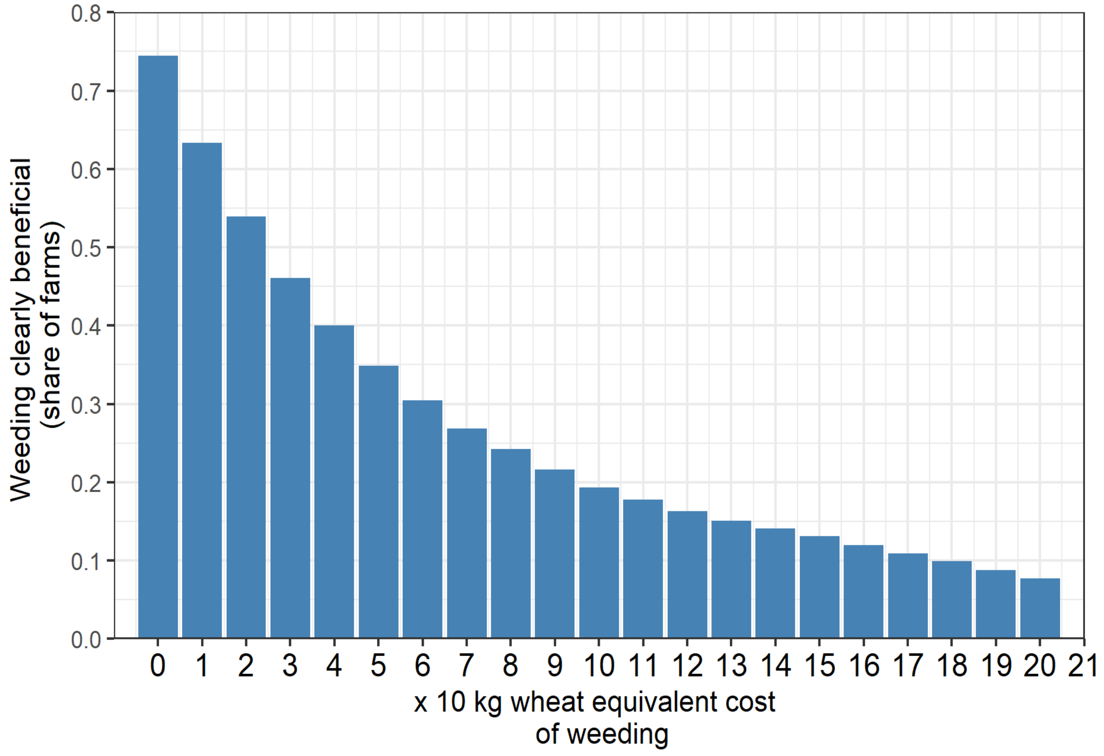
Willingness to pay bounds (ton/ha)

	Upper bound	Lower bound
Mean	0.059	0.052
Standard deviation	0.081	0.080
Minimum	-0.127	-0.133
10th Percentile	-0.011	-0.018
25th Percentile	0.008	0.001
Median	0.034	0.028
75th Percentile	0.084	0.077
90th Percentile	0.189	0.180
Maximum	0.527	0.513
Clearly better (%)	**76**	
Clearly worse (%)	**18.3**	
Number of farms	**8550**	

Benefits of weeding under risk and cost sensitivity: Is weeding beneficial?

Proportion of farmer for which weeding is clearly beneficial even if they were risk averse





CIMMYT

Incorporating behavioural economics in CSA advisories

Some behavioural biases

- Time preference: Farmers as many other agents are present biased.
- Risk preference: Risk averse, loss averse
- Bounded rationality and computational costs

Suggestions for step-wise risk assessment

- Back-end
  - Approximate probabilistic and risk assessments
  - Use more conservative, and more ambitious probabilistic assessments to recommend (if IRR >30% for 70% of the modelled for sowing week X then recommend)
  - Assume a risk neutral=profit assessment, assume risk averse= second order stochastic dominance
- Direct risk question [focus on things correlated with risk preferences like patience, imitation, e.t.c]
  - Are you testing new approaches every year or you always practice what everyone does in the village?
- Direct risk question [Better if visuals are possible, e.g., 2 skys without raindrops, 4 carts 1 tons each, e.t.c)
  - If there is a late monsoon 2 times every 10 years in your village, would you plant a nursery of long duration variety that gives 4 ton yield if monsoon comes on time but only 2 ton if it comes late?
  - If no, what if late monsoon is 4 times every 10 years, would you now plant nursery with long duration variety?
  - If no, what if late monsoon is 8 times every 10 years, would you now plant nursery with long duration variety?
  - If no, what if late monsoon is 10 times every 10 years, would you now plant nursery with long duration variety?

Holistic Assessment

- Profitability and risk assessments are a component of many other dimensions
- Think of these as characteristics of the technology itself (e.g., is cost of transplanting labor/availability the same for early sowing). What is different between Nov 5 and Nov 15?
- A robust dynamic advisory should also consider endowments [i.e., land/labor ratio]
- Technology-farmer characteristics matrix: Simple check would be ordering technology characteristics and endowment characteristics (or induced innovation indicators).
- Think of online recommender systems [movies, amazon, e.t.c]

References

- **Spatial Probabilistic Assessment Tool**
  - Chai, Y., Pardey, P.G., Hurley, T.M., Senay, S.D., and Beddow, J.M. 2020. "A Probabilisitic Bio-Economic Assessment of the Global Consequences of Wheat Leaf Rust." *Phytopathology* 110 (12). Doi: https://doi.org/10.1094/PHYTO-02-20-0032-R.
  - Beddow, J.M. 2015. "Research investment implications of shifts in the global geography of wheat stripe rust." *Nature Plants* 1: 15132. Doi: https://doi.org/10.1038/nplants.2015.132.
  - Chai, Y., Senay, S., Horvath, D., and Pardey, P. 2022. "Multi-peril pathogen risks to global wheat production: A probabilistic loss and investment assessment." *Frontiers in Plant Science* 13: 1034600. Doi: 10.3389/fpls.2022.1034600 .
  - McCullough, E.B., Quinn, J.D., and Simons, A.M. 2022. "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.
- **Spatial Risk Assessment Tool**
  - 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(6): 711-723. Doi: https://doi.org/10.1111/agec.12454.

References

- **Portfolio Optimization and Mean-Variance Toolkit**
  - Barkley, A., Peterson, H.H., and Shroyer, J. 2010. "Wheat varietal selection to maximize returns and minimize risk: an application of portfolio theory." *Journal of Agricultural and Applied Economics* 42(1): 39-55. Doi: https://doi.org/10.1017/S107407080000328X.
- Stochastic dominance
  - Ward et al 2018. Demand for Complementary Financial and Technological Tools for Managing Drought Risk. Economic Development and Cultural Change.
- **Reviews: Analysing price and production risks**
  - Hurley et al 2010. A Review of Agricultural Production Risk in the Developing World
  - Boyd, C., and Bellemare, M.F. 2020. Microeconomics of Agricultural Price Risk. Annual Review of Resource Economics
  - Tack, J., and Yu, J. 2022. "Risk management in agricultural production." Handbook of Agricultural Economics. Doi: https://doi.org/10.1016/bs.hesagr.2021.10.004.

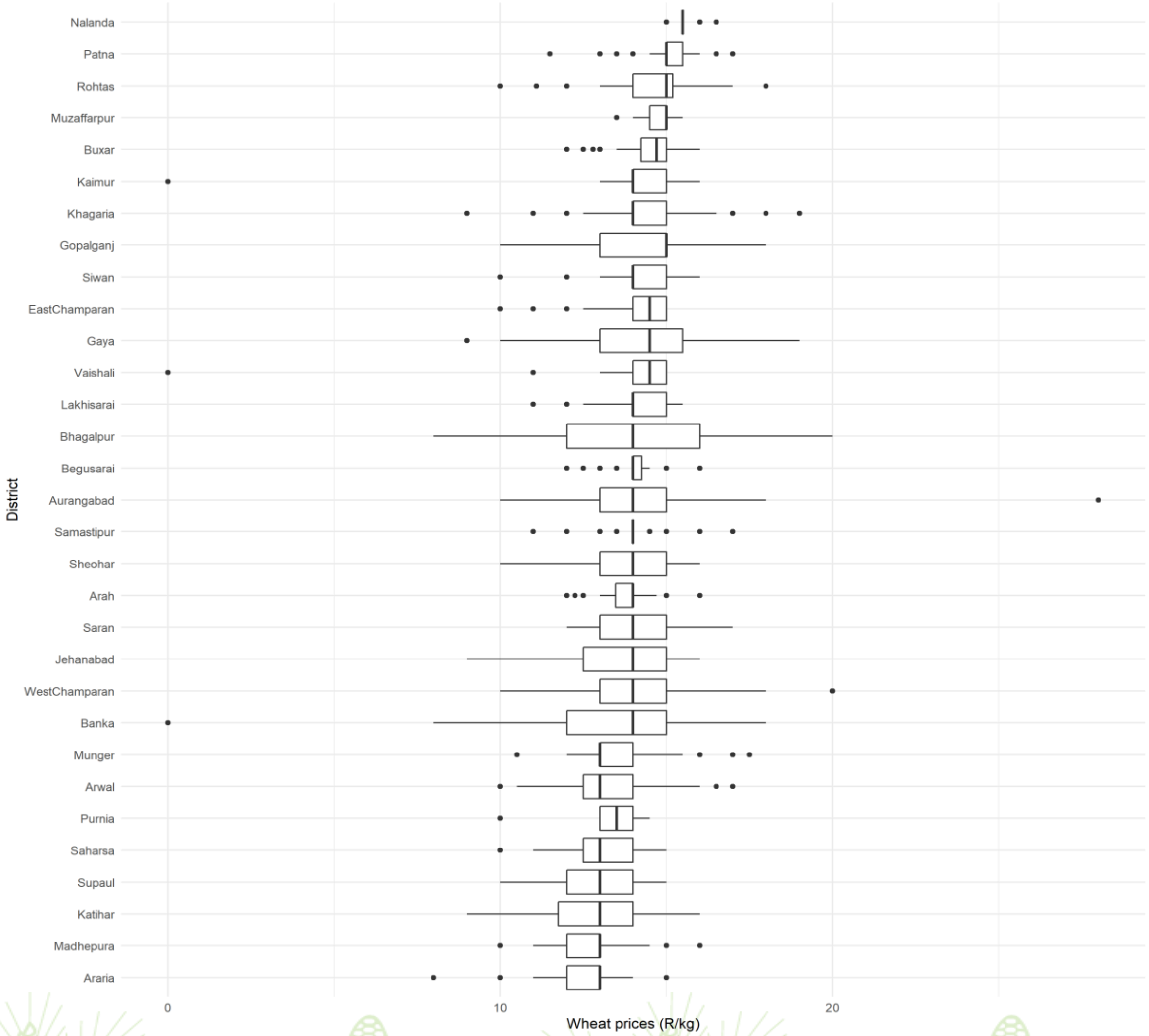
Thank you

Contact: m.mkondiwa@cgiar.org

Appendices

Herbicide application economic assessment

Output price

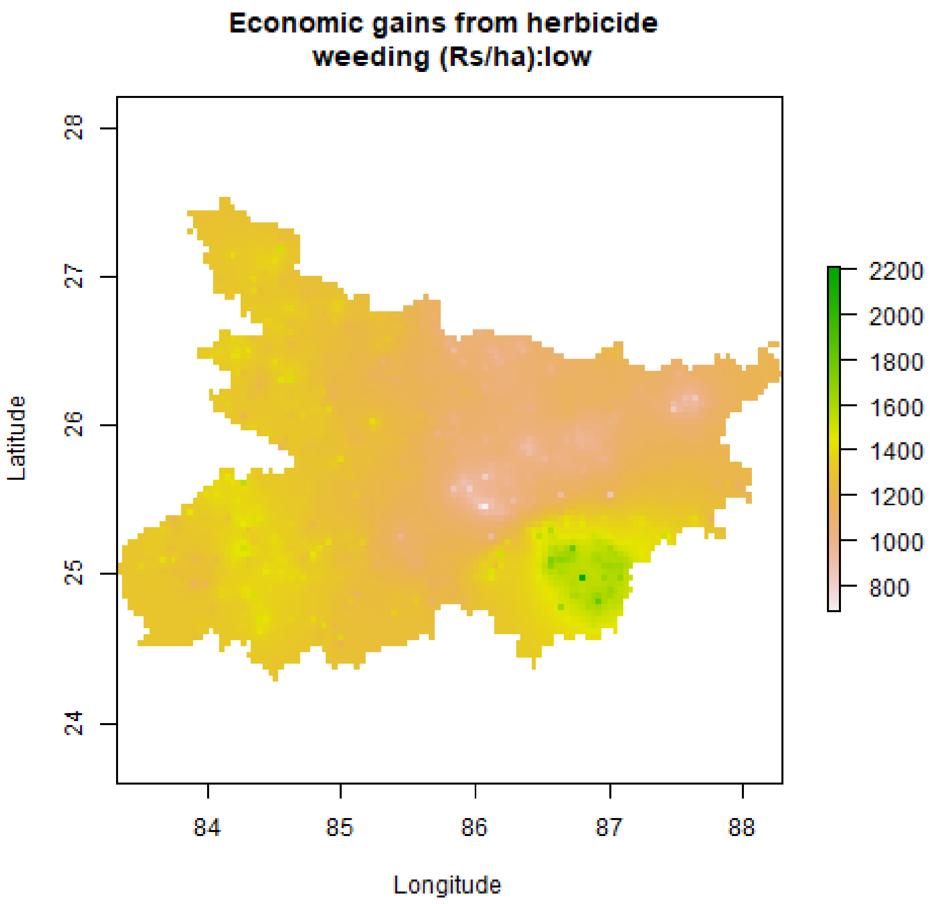


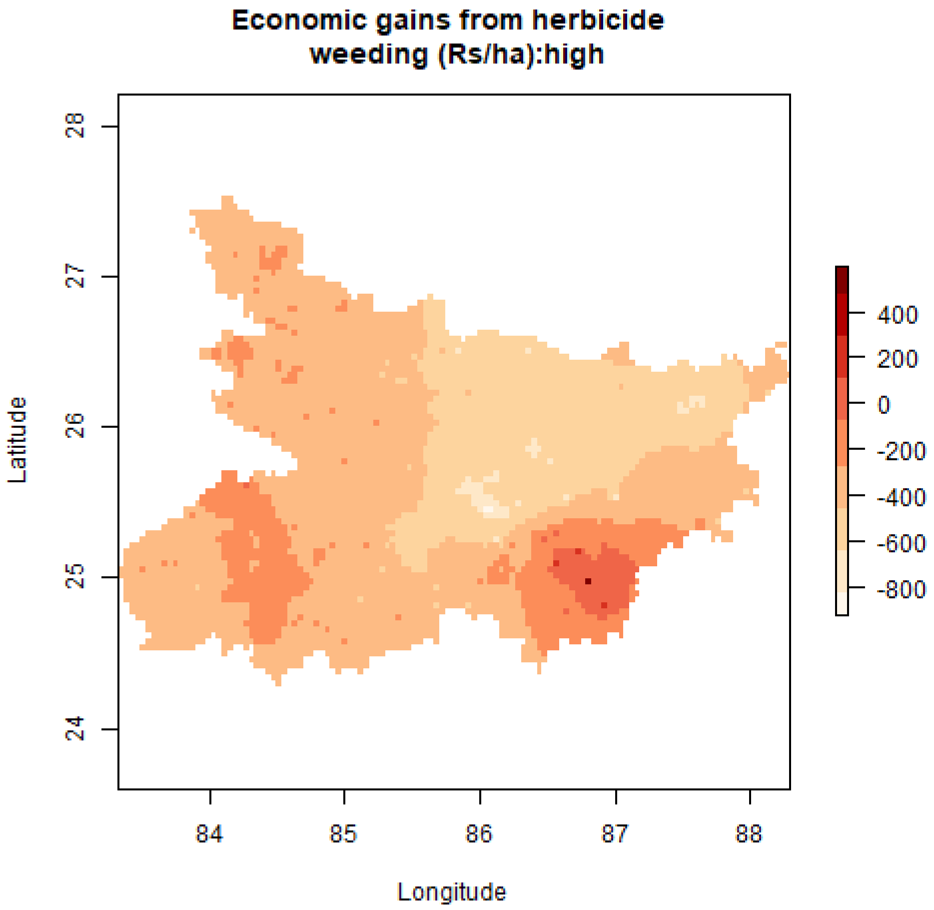
CIMMYT

Prices of major wheat herbicides in India

Herbicide	Weeds killed	Price (Rs/kg)	Price USD/kg (1USD=80 Rs)	Cost per ha based on dosage	USD/ha
24D	Broadleaf	340.00	4.25	340.00	4.25
Sulfosulfuron+ Metsulfuron	Broadleaf +Grassy	12926.00	161.58	517.04	6.46
Sulfosulfuron	Broadleaf +Grassy	3327.00	41.59	109.79	1.37
Clodinaforp	Grassy	642.00	8.03	256.80	3.21
Pendimethalin	Broadleaf +Grassy	395.00	4.94	1659.00	20.74
Carfentrazone	Broadleaf	1025.00	12.81	51.25	0.64
Metsulfuron	Broadleaf	3223.00	40.29	64.46	0.81
Total	Broadleaf	12926.00	161.58	517.04	6.46
Isopruton	Grassy	362.00	4.53	481.46	6.02

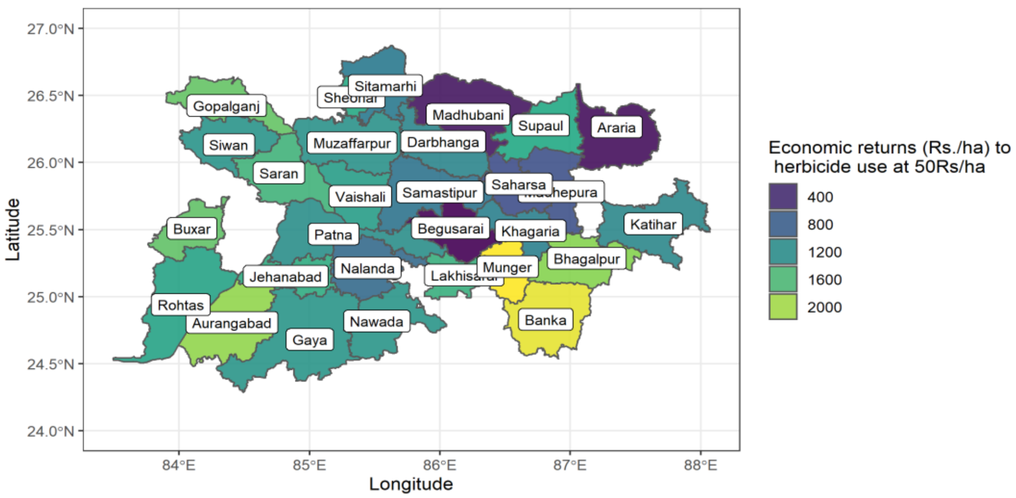
Economic benefits [Herbicide]



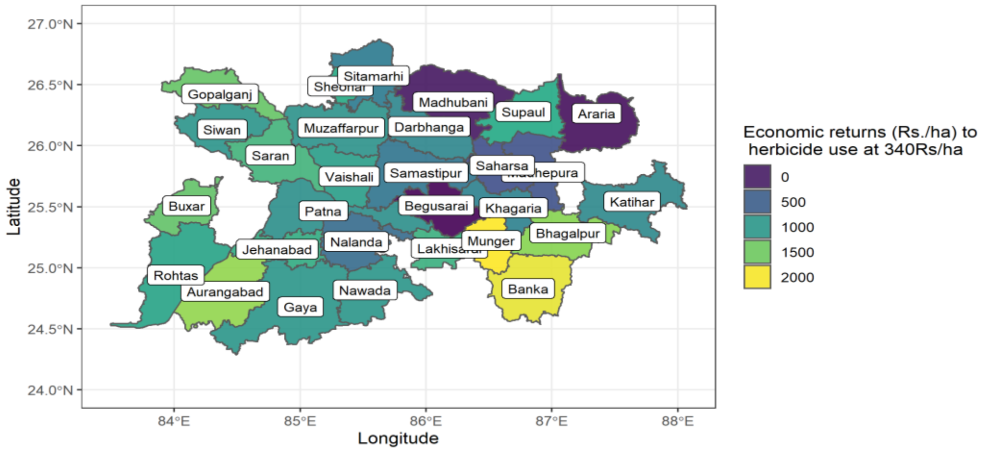


Economic returns

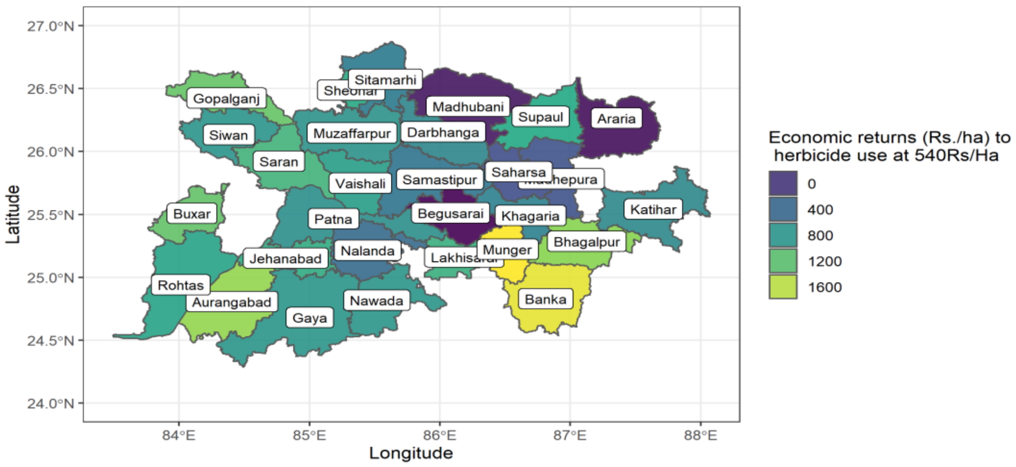
At Carfentrazone price



At 2,4 D price



At Metsufuron +Sulfusoron rice



At Pendimethalin Price