Causal machine learning for fertilizer recommendations: contextual bandit policy improves profit in offline evaluation

by Daniel R. Doran

B.S. in Mathematics, May 2013, University of Southern California
M.S. in Applied Mathematics, December 2016, University of Southern California
Masters in International Development and Cooperation, January 2023, University of Lisbon

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

January 9, 2026

Praxis directed by

Haya Shajaiah
Professor of Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that Daniel Ryan Doran has passed the Final Examination for the degree of Doctor of Engineering as of November 7, 2025. This is the final and approved form of the Praxis.

Causal machine learning for fertilizer recommendations: offline bandit policy improves profit in historical evaluation

Daniel R. Doran

Praxis Research Committee:

Haya Shajaiah, Professorial Lecturer in Engineering and Applied Science, Praxis Director

Amir Etemadi, Adjunct Professor of Engineering and Applied Science, Praxis Chair

Jeffrey Yu, Professorial Lecturer of Engineering Management and Systems Engineering, Committee Member

Dedication

The author wishes to dedicate this research to the farmers, agronomists, and rural communities who steward land and livelihoods. We offer this work in service of evidence-based, sustainable agronomy that can strengthen incomes, resilience, and equity – and to inspire further research that turns data into public-good insight.

Acknowledgements

The author wishes to thank the farmers and extension partners in Chiapas who contributed to the maize dataset, and the International Maize and Wheat Improvement Center (CIMMYT) for open data access.

Abstract of Praxis

Causal machine learning for fertilizer recommendations: contextual bandit policy improves profit in offline evaluation

This praxis develops and validates a causal machine learning framework for optimizing fertilizer recommendations in smallholder maize systems, using historical farm data from Chiapas, Mexico. Smallholder yields in the region remain below potential due to a lack of fertilizer guidelines that take into account heterogeneity in soil conditions, climate, and management. To address this gap, the study formulates fertilizer recommendation as a one-step offline contextual bandit problem, integrating ensemble surrogate crop modeling with conservative policy optimization and robust off-policy evaluation.

A multi-year dataset (2012–2018) comprising 4,585 maize field observations was used to train a stacked ensemble surrogate model – combining XGBoost, LightGBM, and CatBoost learners with a ridge meta-learner – to predict profit responses to nitrogen, phosphorus, and potassium (N–P_2O_5–K_2O) applications under varying conditions. This model achieved an average R^2 of 0.65 and an RMSE of 3,561 MXN/ha on unseen data, demonstrating strong predictive performance in a highly variable agronomic and economic context. The surrogate serves as the foundation for a conservative contextual bandit policy constrained to historically-supported fertilizer regimes, ensuring that its fertilizer recommendations maintain agronomic and empirical plausibility.

Policy performance was evaluated offline using doubly robust (DR) and self-normalized doubly robust (SNDR) estimators. Across the evaluation period, the learned

policy achieved statistically significant profit improvements over the historical baseline while satisfying pre-specified support criteria and exhibiting well-behaved importance weights – confirming reliability of the offline estimates. These results show that conservative, data-driven policies can extract economically meaningful insights from observational agronomic data without new field experiments.

The validated policy was operationalized through a prototype bilingual decision-support web application that translates the model into interpretable, site-specific N–P_2O_5–K_2O recommendations with predicted profit outcomes. This praxis thus bridges the gap between machine learning research and practical agricultural decision support, demonstrating how conservative causal and reinforcement-learning principles can enhance smallholder profitability and input efficiency.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Symbols

x	Context feature vector (soil, pre-plant weather, management, etc.)
a	Action (discretized fertilizer choice on the N–P_2O_5–K_2O grid)
Y	Maize yield
P	Net profit (maize revenue minus fertilizer costs)
Mg	Metric ton
ha	Hectare
MXN	Mexican peso
$\hat{\mu}$	Surrogate profit model
$\hat{v}$	Auxiliary surrogate yield model
π_0	Logging (baseline) policy
π_1	Learned target/recommendation policy
$\pi_{greedy}(x)$	Deterministic greedy policy produced by the surrogate model
$\hat{V}(\pi)$	Estimated profit value of policy π (MXN/ha)
A	Action set
$\Delta(A)$	Set of all distributions over the action set A
$\delta(\cdot)$	Dirac distribution
$g(x)$	Greedy supported action for context x
$S(x)$	Set of supported actions for context x
$U_{support}$	Uniform distribution over supported actions
w_i	Importance weight (density ratio) for sample i
λ	Baseline-mixing coefficient

ε	Exploration rate in ε-greedy on-support exploration
$\hat{k}$	Estimate of GPD tail shape parameter

List of Acronyms

ANOVA	Analysis of Variance
AI	Artificial Intelligence
APSIM	Agricultural Production Systems sIMulator
CatBoost	Categorical Boosting
CATE	Conditional Average Treatment Effect
CEC	Cation Exchange Capacity
CI	Confidence Interval
CIMMYT	International Maize and Wheat Improvement Center
CV	Cross-Validation
DOY	Day of Year
DM	Direct Method
DR	Doubly Robust
DRos	Doubly Robust with optimistic shrinkage
DSSAT	Decision Support System for Agrotechnology Transfer
ECE	Expected Calibration Error
ESS	Effective Sample Size
GHG	Greenhouse Gas
GPD	Generalized Pareto Distribution
IDE	Integrated Development Environment
IID	Independent and Identically Distributed
IPCC	Intergovernmental Panel on Climate Change

IPS	Inverse Propensity Scoring
IS	Importance Sampling
K	Potassium (K_2O)
LAI	Leaf Area Index
LCB	Lower Confidence Bound
LightGBM	Light Gradient Boosting Machine
LFBR	Local Blanket Fertilizer Recommendations
MDP	Markov Devision Process
MI	Mutual Information
ML	Machine Learning
N	Nitrogen
NBFR	National Blanket Fertilizer Recommendations
NDVI	Normalized Difference Vegetation Index
NUE	Nitrogen Use Efficiency
NuUE	Nutrient Use Efficiency
OPE	Off-Policy Evaluation
P	Phosphorus (P_2O_5)
PSIS	Pareto-Smoothed Importance Sampling
QUEFTS	QUantitative Evaluation of the Fertility of Tropical Soils
R^2	Coefficient of Determination
RCT	Randomized Control Trial
RL	Reinforcement Learning
RMSE	Root Mean Square Error

SMS	Short Message Service
SOM	Soil Organic Matter
SNDR	Self-Normalized Doubly Robust
SNIPS	Self-Normalized Inverse Propensity Scoring
SPIBB	Safe Policy Improvement with Baseline Bootstrapping
SSNM	Site-Specific Nutrient Management
STFF	Soil Testing and Formula Fertilization
Switch-DR	Switch Doubly Robust
VRN	Variable Rate Nitrogen
WIS	Weighted Importance Sampling
WUE	Water Use Efficiency
XGBoost	Extreme Gradient Boosting
YH	Yield History

Chapter 1—Introduction

1.1 Background

Maize is a staple crop in Chiapas, Mexico, cultivated predominantly by smallholder farmers using traditional practices. Average maize yields in the region sit around 1.9 Mg/ha statewide – with more productive sub-regions like Frailesca averaging 3.5 Mg/ha (Martinez et al., 2020). This heterogeneity suggests that some farms are certainly below their potential attainable levels, indicating a clear yield gap. There are many reasons for this gap: many farmers continue to plant maize following traditional *milpa* cropping methods, for example, which minimize external inputs (including fertilizer). One study found negligible yield gains from adding fertilizer to maize crops in Chiapas (Fonteyne et al., 2022). Meanwhile, other studies in the region do show positive yield responses to fertilizer inputs (Trevisan et al., 2022), suggesting that Chiapas could benefit from tailored fertilizer management. Beyond the yield gap, there is the arguably more important profitability gap; even where yields might be increased, it must be done in a cost-effective manner. If adding more fertilizer increases yield, but without making up for the added fertilizer costs, the extra expense may not pay off for farmers. This is a decisive consideration, as smallholder farmers typically operate under tight economic constraints. Finally, there are environmental considerations; improving fertilizer management is thus not just about raising yields, but also about maximizing input efficiency, and doing so sustainably.

Agricultural extension services – defined by Birner et al. (2009) as "the entire set of organizations that support and facilitate people engaged in agricultural production to

solve problems and to obtain information, skills, and technologies to improve their livelihoods and well-being" (p. ix) – often rely on blanket regional or national fertilizer recommendations. These kinds of homogeneous recommendations generally overlook local variation in soil, weather, and crop needs (Chernet et al. 2024), which leads to inefficiencies. Some farmers over-apply fertilizer, over-spending and harming the environment, while others under-apply fertilizer and give up potentially profitable yield (Vallepogu et al., 2024). This inefficiency is readily observed in practice in smallholder systems: Chernet et al. (2024) document, for example, that blanket fertilizer recommendations in Sub-Saharan Africa have resulted in suboptimal profitability and yield outcomes, and demonstrate the utility of site-specific guidance (Chernet et al., 2024). In Mexico as well, as is the case in many developing regions, smallholder farmers often lack access to advanced technology or decision support, opting instead to rely on generalized national or regional guidance.

In recent years, data-driven techniques in agriculture has emerged as a promising direction to provide more tailored and precise agronomic advice. The proliferation of farm data and advancement of artificial intelligence (AI) and machine learning (ML) techniques give rise to the development of models that can predict localized crop responses, such as crop yield, and optimize inputs, such as nutrient or water usage. Numerous studies have shown that data-driven, site-specific nutrient management (SSNM) can significantly outperform traditional farming practices. For instance, Chernet et al. (2024) use machine learning to generate SSNM recommendations for wheat in Ethiopia, achieving a 16 and 25% higher yield output (and a 33% and 19% profit uplift) respectively, compared to local- and national-level recommendations from advisory

services. Similarly, a meta-analysis of data-driven SSNM programs across Asia and Africa (Chivenge et al., 2021) established that site-specific fertilizer advice increased yields by 12%, and profits by 15%, on average, while reducing nitrogen usage by 10%. Results such as these highlight the potential of data-driven agricultural support in providing greater crop yields and profits for farmers and minimizing environmental impact.

1.2 Research Motivation

A comprehensive dataset covering seven seasons of maize field trials in Chiapas (2012-2018), published by Trevisan et al. (2022), offers a unique case study for analyzing in a data-driven way how maize yields respond to differing management and environmental conditions. This dataset includes diverse agro-environments, across different elevations and microclimates of Chiapas, a range of soil properties, differing weather patterns, along with records of farmers' management such as fertilizer rates and planting practices. Such historical data can be leveraged to build a surrogate crop profit model – a machine learning model that predicts net profit (calculated as maize revenue minus fertilizer costs) given a context (location, soil, weather) and an action (fertilizer application rate). Surrogate models have been widely used to emulate complex agricultural systems when direct experimentation is infeasible (Corrales et al., 2022; Cunha et al., 2023). By training a surrogate on field data, researchers can simulate the outcome of untested input decisions with a reasonable degree of confidence, effectively creating a virtual test-bed for optimization.

Reinforcement learning (RL) offers a complementary approach to exploit such a surrogate model for decision optimization. Offline RL – learning policies from previously collected (offline) data rather than through online, trial-and-error interaction – is particularly relevant for agriculture, where conducting new trials is costly and time-consuming. Offline RL algorithms allow us to derive an optimal policy from a static dataset without additional field experimentation. Given the dataset in question, a one-step offline contextual bandit formulation is well-suited for fertilizer recommendation: each field-season can be treated as a one-shot decision problem where the algorithm observes a context (soil type, rainfall pattern forecast, etc.) and must choose an action (fertilizer dose) to maximize expected reward (profit). Unlike a full sequential RL problem, there is no long horizon of repeated actions – but the decision is contextualized by environment conditions. Contextual bandit algorithms have seen success in real-world applications like personalized recommendations, ad placement, and healthcare, by learning decision policies that adapt to situational features. In agriculture, researchers are beginning to explore bandit and RL methods for management recommendations. For example, Moothedath et al. (2023) proposed a contextual bandit framework to recommend crop varieties, fertilizer, and irrigation levels based on farm-specific context (soil and weather), aiming to maximize profit in maize farming. Their strategy, which respects farmers' pre-specified constraints, underscores the importance of conservatism and appropriate guardrails in agricultural decision support.

Part of the motivation for this study thus arises from a practical need for trustworthy recommendations. Many machine learning-based methods risk overfitting and suggesting fertilizer regimes that are not agronomically, environmentally, or

economically plausible. Farmers and extension agents require assurances that any advice is safe to implement in practice and interpretable, which is where causal inference methods can really shine. Thus, this praxis focusing on deriving conservative, empirically-grounded fertilizer recommendations – leveraging machine learning predictive models and strict historical support criteria to ensure any proposed recommendation demonstrably outperforms farmers' historical practice only when the evidence is statistically sound.

Taken together, the rich agronomic Chiapas dataset, modern machine learning crop modeling, an offline contextual bandit policy framework, and rigorous policy evaluation using causal inference methods provides a unique opportunity to develop a novel fertilizer recommendation solution for smallholder farmers in the region. This study learns and presents a conservative fertilizer policy, that remains strictly within historically-supported fertilizer regimes, aimed at optimizing farmer's net profits. The policy is learned offline, utilizing solely the historical Chiapas farm data (Trevisan et al., 2022).

Finally, this praxis is motivated by the broader goal of translating advanced analytical techniques into accessible decision support. The study develops a bilingual Streamlit web application to showcase the policy in an easy-to-use, site-specific fertilizer recommendation tool, with the knowledge that on-farm validation trials should precede any real-world deployment of such a policy, and that recommendations from such a policy could be made even more accessible if delivered to farmers through channels like WhatsApp or SMS messages. This praxis lies, therefore, at the confluence of agricultural

development, machine learning, and causal inference, and aims to encourage more equitable access to data-driven techniques to rural communities.

1.3 Problem Statement

Blanket fertilizer recommendations allocate resources suboptimally given field heterogeneity; maize effectively utilizes only 40% of applied fertilizer on average (Quan et al., 2021), leaving yield and profit gains on the table, and increasing waste.

Smallholder maize farmers in Chiapas, Mexico lack site-specific, evidence-based guidance on fertilizer management, leading to scenarios where farmers under-apply fertilizer, leading to yield gaps, or over-apply it, unnecessarily spending financial resources and causing harm to the environment. With a finite capacity of extension services in the region, and acknowledging the resources required for large-scale on-farm trials to refine recommendations, many farmers likely remain exposed to these fertilizer inefficiencies.

Under-application of fertilizer means that farmers forgo otherwise attainable yield and crop revenue. On the other hand, over-application of fertilizer is particularly pernicious, since using more fertilizer than necessary harms the environment and at the same time reduces profits for farmers via extraneous fertilizer costs. The core question, therefore, is how to generate trustworthy fertilizer recommendations that account for site-specific conditions to optimize farmers' profits, and minimize waste in the process? A solution for this is presented in the following section.

1.4 Thesis Statement

A fertilizer recommendation tool, trained and evaluated on historical farm data, will increase farmers' net profits while staying within historically-observed fertilizer application regimes.

This praxis asserts that a tool utilizing an offline contextual bandit fertilizer policy, built upon a machine learning model to predict farmers' net profits and evaluated using causal inference techniques, can generate context-specific fertilizer recommendations to meaningfully improve profits. It is ensured that recommendations are trustworthy by proposing fertilizer combinations only when there is sufficient evidence for their success from historical farm trials, thereby avoiding extrapolations and proposing unrealistic combinations.

1.5 Research Objectives

To address the stated problem, the research pursues the following objectives:

- **Objective 1: Train and validate an ensemble surrogate model.**

  Fit a stacked tree-based surrogate (XGBoost/LightGBM/CatBoost with ridge meta-learner) to predict net profit (defined as crop revenue minus fertilizer costs) as a function of context and N–P_2O_5–K_2O actions; evaluate with forward-by-year cross validation (CV) and report RMSE and R^2 for out-of-fold predictions. Fit an identical surrogate to predict yield to isolate agronomic response from price effects.

- **Objective 2: Train a joint propensity model (π_0) for the action space.**

  Estimate calibrated propensities over a discretized N–P_2O_5–K_2O actions grid; verify action-space coverage and reliability to enable stable importance weighting and support checks.

- **Objective 3: Learn a conservative new policy (π_1) that achieves greater net profits than farmer's historical practice (π_0).**

  Optimize a contextual-bandit policy constrained to historically supported action cells, with overlap and baseline-mixing safeguards to minimize extrapolation beyond observed fertilizer regimes.

- **Objective 4: Perform rigorous offline evaluation of the new policy.**

  Assess π_1 against historical farmer practice with doubly-robust (DR) and self-normalized doubly-robust (SNDR) estimators, cluster-bootstrap CIs, and pre-specified support criteria with well-behaved importance weights.

- **Objective 5: Service an interpretable advisory app.**

  Operationalize the validated policy in a farmer- and extension-facing web app that accepts local conditions and returns supported N–P_2O_5–K_2O recommendations with expected outcomes and confidence flags, linking the research to actionable advice for smallholders and partners.

Through these objectives, the study will produce a validated framework for contextual-bandit fertilizer optimization and lay the groundwork for integrating this framework into an accessible decision support tool.

1.6 Research Questions and Hypotheses

This research is guided by three primary research questions (RQ) and their associated hypotheses (H). These questions address the feasibility and effectiveness of combining ML surrogate modeling with offline evaluation for fertilizer optimization, as well as the system's capacity to provide profitable recommendations.

- **RQ1. Predictive performance:**

  How accurately can the stacked ensemble surrogate model estimate profit in diverse smallholder contexts under forward-by-year validation?

- **H1:** The stacked ensemble will achieve strong out-of-fold performance – reflected in an R^2 of at least 0.60 and a RMSE of at most 4,000 MXN/ha – under the specified cross-validation procedure.
- **RQ2. Overlap and propensity reliability:**

  Are behavior propensities $\pi_0(a \mid x)$ well-calibrated and is evaluation overlap adequate on the discretized action grid?
- **H2:** With calibration enabled where class counts allow, the joint propensity model will exhibit a test Expected Calibration Error (ECE) of less than 0.05 with reliability curves close to the 45° line, and evaluation overlap covering at or above 33 of the 36 action cells.
- **RQ3. Decision quality and robustness:**

  Do support-constrained recommendations improve estimated outcomes relative to baselines while maintaining stability?

- **H3:** The constructed π_1 will show positive SNDR/DR uplift over a masked π_0, with cluster-bootstrap CIs satisfying pre-set acceptance criteria, and an overall percent profit uplift of 5% over farmer's historical practice.

1.7 Scope of Research

This praxis is confined to the context of rainfed maize systems in Chiapas, Mexico, and specifically to learning and evaluating nitrogen, phosphorus, and potassium fertilizer recommendation policies. The research leverages an existing dataset of maize field observations collected in Chiapas during 2012-2018; thus, all modeling and policy derivation occur within the scope of this dataset's feature space (soil properties, climate variables, fertilizer rates) during these years. The surrogate model will be trained and evaluated on this dataset, using cross-validation and other techniques, and the contextual bandit policy will be learned offline using these historical examples.

The focus is on *in silico* performance assessment via off-policy evaluation of the policy compared to historical farmer practice, and a prototype decision support tool that operationalizes the learned policy for demonstration is developed. By design, the system does not recommend actions beyond the data's empirical support. This study does not conduct online learning or deploy the learned policy on real farms; field implementation is discussed conceptually but is outside the immediate scope of this praxis. Finally, external validity outside Chiapas is not claimed without retraining on local data.

1.8 Research Limitations

Dependence on off-policy evaluation (OPE) and model fit. Several limitations stem from the chosen scope and methodology. First, as this study does not run new field trials, policy value is estimated from historical data via doubly robust (DR) and self-normalized doubly robust (SNDR) estimators. While DR and SNDR reduce bias and are standard in contextual bandits, they still rely on correct specification of either the outcome model or the behavior policy and can degrade when importance weights concentrate (low effective sample size). The proposed pipeline mitigates these risks with conservative support criteria and requiring well-behaved importance weights; nonetheless, OPE remains an approximation and real-world trials are ultimately needed to verify gains.

Limited extrapolation and support constraints. To avoid evaluating recommendations outside the data's support, the study restricts π_1 to historically observed N–P–K regimes and enforce overlap thresholds. This design trades coverage for credibility: it protects against optimistic bias from extrapolating the surrogate beyond supported actions, but it also means some potentially beneficial recommendations are intentionally out of scope. The tension between overlap requirements and policy ambition is well documented in the OPE literature (Sachdeva et al., 2020).

Observational data and confounding. The dataset reflects farmers' historical choices, not a randomized experiment. Unmeasured management decisions, such as allocating higher inputs to better-managed or higher-potential fields, may confound associations between fertilizer and yield. Although DR estimators help when either the outcome or behavior model is correct, unmeasured confounders can still bias both

modeling and evaluation; causal identification with observational data assumes fertilizer choices are as-if random with respect to potential outcomes.

Price assumptions. The surrogate profit model constructs its reward function assuming fixed prices (N at 16 MXN/kg, P_2O_5 at 12 MXN/kg, K_2O at 8 MXN/kg, and maize at 3.5 MXN/kg) in all training and evaluation years, due to difficulty obtaining precise price data during this period. These prices were informed by *inegi.org.mx*. In reality, fertilizer and grain prices fluctuate year-to-year and throughout the year.

Non-stationarity and external validity. While the surrogate model pipeline is temporally aware – it is trained and evaluated in a forward-by-year manner, re-fitting per evaluation year with recency-weighted training and including a *Year* covariate – the learned policy is offline and static, derived from 2012–2018 data. Climate change or evolving farmer practices could shift the environment in the future, and without online adaptation, the policy might become suboptimal. This praxis does not implement an online learning update due to the offline scope, but acknowledges that periodic re-training with new data would be necessary in practice. Additionally, the study's geographical focus in Chiapas means the model is specialized to that region's agro-ecological conditions; its direct transferability to other regions or cropping systems would require caution and likely retraining on local data.

Practical deployment factors. On the practical side, while the praxis implements a web app that turns the validated policy into interpretable N–P–K recommendations, and while it discusses the potential deployment via WhatsApp or SMS to reach a broader audience, it does not examine rigorously how the web application would be deployed in practice nor delve into the specifics of connecting to the aforementioned communication

tools. It assumes such deployment is feasible, but the actual adoption by farmers is outside the scope. For instance, factors like digital literacy, trust in automated recommendations, and economic capacity to buy the recommended fertilizer amount are all important considerations that are not directly studied in this praxis. Even if the proposed technical solution is sound, these socio-economic considerations could limit real-world impact if not addressed with care.

Ultimately, this praxis is an offline study. Its contributions lie in combining machine learning methods with casual inference techniques in a novel way to promote sustainable agricultural development. The findings must be interpreted with the understanding that on-farm trials would be needed to validate outcomes, and to understand how farmers and extension agents would engage with the recommendations in practice. Despite these limitations, this work provides a crucial stepping-stone for further research by demonstrating the potential of these techniques to support agricultural solutions.

1.9 Organization of Praxis

This praxis is organized into 5 chapters. Chapter 1 has introduced the context, problem statement, and the proposed solution, and outlined the objectives, research questions, and hypotheses guiding the study. Chapter 2 presents a comprehensive literature review, situating the research in the current body of knowledge. Chapter 3 details the research methodology, from describing the dataset used, to training and evaluating the surrogate profit model, constructing the contextual bandit fertilizer policy,

and offline policy evaluation (OPE). Chapter 4 will report the results in detail, to include the performance results of the predictive surrogate model, the expected profit and yield increases from the learned fertilizer policy, as well as expected fertilizer changes. Lastly, Chapter 5 will discuss the findings and analyze them in light of the research objectives, examine their contribution to the body of knowledge, and recommend avenues for further study.

Chapter 2—Literature Review

2.1 Agronomic Decision Support – Fertilizer Recommendations

Agronomic decision support systems are designed to assist farmers in making more informed management choices, such as determining optimal fertilizer or water application rates to improve crop yields or resource efficiency. Utilizing fertilizer is known to substantially improve crop productivity, increasing yield gains by roughly 30-50% (Stewart et al., 2005), though misuse can lead to environmental degradation, economic losses, or even yield declines.

Soil Testing and Formula Fertilization (STFF) is a fertilizer recommendation method popular in China that utilizes regressions based on soil tests and crop yields and is grounded in agronomic principles like Liebig's "minimum nutrient law" and the law of diminishing returns (Gao et al., 2023). Gao et al. (2023) notes that a "unified and precise formula" fails to capture the yield-fertilizer-soil relationship, since in practice small-scale farms exhibit significant heterogeneity in soil conditions and management practices. This limitation has made it difficult to implement traditional STFF at scale and accentuates the need for more adaptive and context-specific agronomic decision support.

In recent years, many such decision support tools have emerged to provide site-specific nutrient management (SSNM) recommendations. One noteworthy example is the Nutrient Expert decision support system (see screenshot in Figure 2-1). This tool requests users to input information on the farmers' management practices, available fertilizer blends, field history, and local conditions and generates tailored fertilizer

recommendations (Orchardson, 2019). The system was calibrated with fertilizer omission field trials in Ethiopia, Tanzania, and Nigeria to estimate fertilizer impacts on yield and refine its advice (Orchardson, 2019). Orchardson (2019) remarks that field evaluations testing the Nutrient Expert tool in Ethiopia demonstrated increased yields, improved fertilizer-use efficiency, and increased profits when farmers followed the tool's recommendations – saving farmers $80/ha on average in input costs.

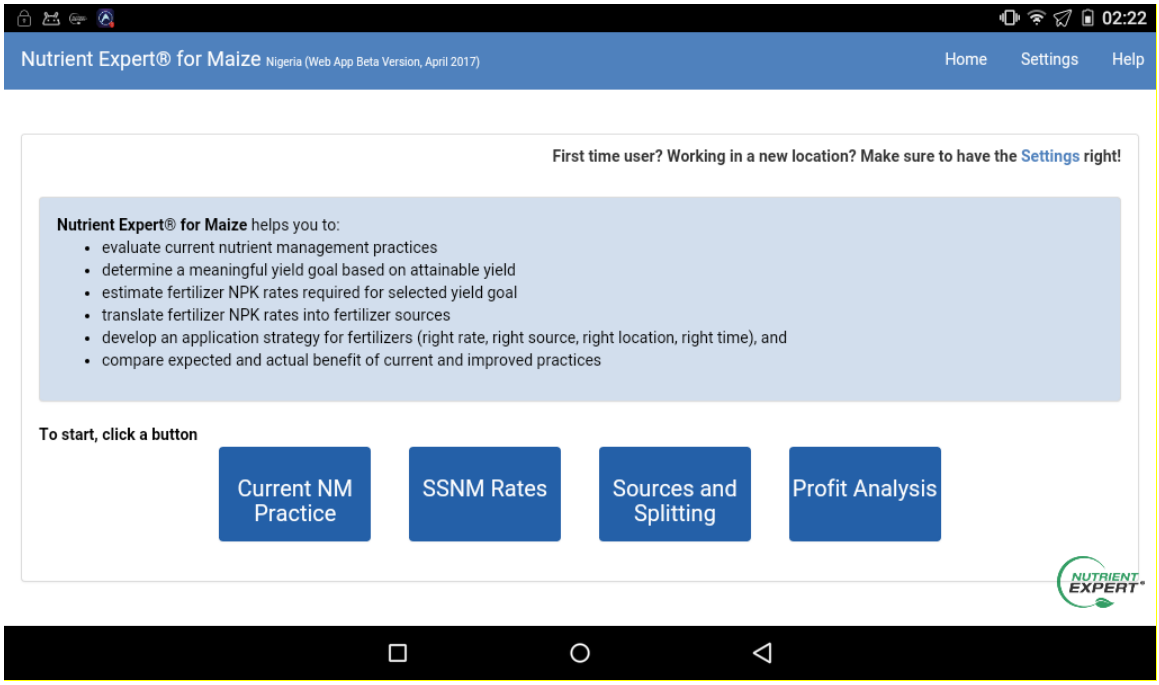

Figure 2-1. Screenshot of the Nutrient Expert tool user interface (Yasabu, 2021)

SSNM decision support is not new – it emerged in the 1990s as a precision farming approach for small-scale cereal production systems in Asia (Chivenge et al., 2021). These systems typically use field-specific data gathered from farmer surveys or

farm trials to calibrate fertilizer advice. Over the last few decades, SSNM decision support tools have been developed for a myriad of crops, such as wheat, rice, and maize, in several geographic contexts. A meta-analysis in Asia and Africa (Chivenge et al., 2021) reports significant benefits of using SSNM: on average increasing yields by 12% and profitability by 15% compared to farmers' historical practices, while reducing by 10% the amount of nitrogen fertilizer applied. These improvements are attributed to a better context-specific matching of fertilizer supply with crop needs among heterogenous field conditions.

Taking another example, Chernet et al. (2024) corroborate these advantages demonstrated from SSNM practices. They train machine learning (ML) models on a dataset of wheat trials in Ethiopia to generate site-specific fertilizer recommendations, which are then validated in further field trials. In these trials across 277 locations, these ML-based recommendations outperformed both the national (NBFR) and local blanket fertilizer recommendations (LBFR) (Chernet et al., 2024). Crop yields increased by 16 and 25%, and profits increased by 19 and 33% respectively; furthermore, nitrogen use efficiency (NUE) rose by 30% (Chernet et al., 2024). This is a notable example of SSNM using modern algorithms being validated through field trials to demonstrate significant end results for farmers.

Therefore, beyond rule-based SSNM decision support, researchers have begun to effectively integrate machine learning and artificial intelligence (AI) into these tools. Machine learning can leverage large agronomic datasets to capture nonlinear interactions between fertilizer, local farm conditions, and crop yield, and thus can model these complex relationships more accurately than other quantitative methods like simple

regressions. Emerging studies have employed reinforcement learning (RL), a branch of ML that trains an agent to make decisions in an environment to optimize one or more rewards, to directly optimize fertilizer management policies. Gautron et al. (2022), for instance, developed a Gym-DSSAT environment, built on a process-based crop model, to train RL agents for enhancing fertilization efficiency in multiple crops. These approaches can iteratively learn from data via simulations, offering adaptive recommendations that potentially outperform static guidelines.

Hence, many AI-driven approaches so far have focused on predicting crop growth, or even optimizing fertilizer recommendations via crop model simulations; however, few systems optimize decisions like fertilizer rates in a way that explicitly maximizes farmer outcomes using causal inference techniques – estimating empirical effects – as is done with field experiments but performed offline with observational data. The present research is motivated by this gap – building upon the successes of agronomic decision tools and ML techniques to develop a methodology that directly links fertilization decisions with causal profit improvements.

The next sections review the relevant literature on machine learning for crop modeling, reinforcement learning approaches to decision-making, and the causal inference and evaluation techniques that will help bridge data-driven models with effective agronomic recommendations.

2.2 Machine Learning in Crop Modeling

Machine learning has become a cornerstone of modern crop modeling and precision agriculture, offering new capabilities for prediction and decision support. Unlike traditional process-based crop models (i.e., DSSAT, APSIM) which use fixed equations to simulate crop growth, ML methods learn patterns directly from data, allowing them to capture complex, nonlinear relationships between inputs (weather, soil, management) and outputs (yield, biomass, profit) (Shahhosseini et al., 2020). This data-driven nature means ML can automatically discover interactions and feature importance that might be unknown or too complicated to encode in mechanistic models. For instance, in agronomic applications, an ML model can be trained on farm data to learn how rainfall, fertilizer usage, and soil conditions interact to affect crop yield (or profits, if prices are also included), instead of assuming a predefined rule-based form. Research has shown that ML algorithms, ranging from polynomial regressions and decision trees, to deep neural networks and advanced ensemble methods, can demonstrate high accuracy in predicting crop yield (Khaki et al., 2019; Shahhosseini et al., 2020). These models excel at handling large datasets and high-dimensional inputs, which are typically present in agronomic use cases, and automatically selecting and weighting the most relevant factors to predict the target variable.

An important advantage of using ML in crop modeling is its ability to integrate data from various sources, including from remote sensing. Research has increasingly made use of data collected from satellites or drones, as well as handheld devices such as multispectral radiometers, to monitor crops (often in near real-time) and predict yields. Ko et al. (2024) explores various ML algorithms (i.e., ridge regression, support vector

regression (SVR), random forest (RF), among others) to model the relationship between vegetation indices and the leaf area index (LAI) of soybean and rice. The vegetation indices were derived from canopy reflectance measurements taken from handheld radiometric sensors, and the LAI were measured using similarly using handheld sensors (Ko et al., 2024). By assimilating the ML-estimated LAI values into a rule-based crop model, they effectively reproduce seasonal crop development under various nitrogen fertilizer treatments, outperforming a Bayesian regression baseline (Ko et al., 2024). This study illustrates how machine learning can provide serve as a bridge between rule-based models and sensor data and thus enhance simulation fidelity. In this way, ML offers nonlinear predictive power and adaptability, whereas rule-based methods contribute mechanistic insights from domain knowledge, as well as the ability to extrapolate beyond the observed data.

Recent studies has demonstrated indeed that incorporating both types of approaches can outperform either method used alone (Droutsas et al., 2022). Shahhosseini et al. (2021), for example, found that integrating simulated outputs from a rules-based APSIM crop model as features in ML models reduced maize yield root mean squared error (RMSE) by 7-20% in the US Corn Belt. The APSIM crop model produced agronomically meaningful features, like soil moisture and water table depth, that the ML algorithm then used to make more informed crop yield forecasts (Shahhosseini et al. 2021). Of note, among the most important predictors that the ML model identified in predicting yield were simulated drought stress and soil hydrology features, which indicates that the APSIM model, through its process-based understanding of agronomic relationships, meaningfully enriches the feature set.

Machine learning models can, additionally, compensate for the shortcomings of more traditional models. Process-based simulations typically require expert parametrization and thorough calibration, and they can easily diverge from realistic outcomes if any process is misspecified (Ko et al., 2024). On the other hand, ML algorithms are relatively easy to implement, require no tuning of domain-specific equations, and can be quickly retrained as new data becomes available. They can run faster in many cases, facilitating rapid iterations and evaluations. As Shahhosseini et al. (2021) point out, machine learning does not require the practitioner to have deep domain expertise in agronomy – the algorithm learns the relationship between inputs and outputs more or less on its own – while running a rule-based crop model demands careful setup of parameters for each scenario tested.

Machine learning in crop modeling has greatly improved researchers' ability to predict crop yields, resources, and diagnose crop health issues. Applications are varied and range from within-season yield forecasts using weather updates and remote sensing to long-term projections under climate change. The logical next step, which will be discussed in the next section, is to integrate ML-based predictions with algorithms that can choose optimal fertilizer actions rather than just predict outcomes.

2.3 Offline Reinforcement Learning and Contextual Bandits

While predictive modeling can help anticipate crop outcomes, another approach is to directly learn policies for decision-making using reinforcement learning (RL). Reinforcement learning, as discussed above, is a machine learning framework in which

an agent interacts with an environment and learns a policy (i.e., a mapping from states or observations to actions) that maximizes a cumulative reward. Classical RL algorithms, such as Q-learning, require active exploration. Via these algorithms, an agent attempts different actions and observes rewards over many iterations to discover strategies that perform well. In agriculture, one can conceptualize using RL for sequential decisions like irrigation scheduling or fertilizer management, where each season or even distinct growth stages within seasons provide feedback in terms of yield or profit. Real-world experimentation is expensive and risky, however, since a farmer or extension agent cannot explore a wide range of strategies on actual fields without incurring real potential yield losses or costs. This is where offline reinforcement learning and contextual bandits become relevant.

Offline RL (sometimes called batch RL) learns policies from pre-collected historical data of past interactions, rather than live experimentation (Che, 2025). For example, one might have historical data of farming decisions, conditions, and outcomes, such as past fertilizer rates, soil conditions, and crop yields. The goal is to learn a new set of fertilizer actions given a specific context that would, if it had been followed instead of the historical fertilizer choices that were made, yield higher rewards. This approach is especially valuable in domains where deploying an untested policy is costly or dangerous. It has gained increasing attention in fields like healthcare, autonomous driving, and recommender systems (Che, 2025). Agriculture shares similar constraints – experiments are time-consuming and can jeopardize livelihoods – so an offline RL approach allows leveraging existing farm data to improve decisions while minimizing on-farm trial-and-error.

A related but simpler framework is the contextual bandit problem. Contextual bandits can be seen as a one-step decision RL: for each independent scenario (context), the agent picks an action and receives a reward, with no long-term state dynamics. In farming, a context could be the features of a specific field-season, like soil properties, weather forecast, and crop variety, and the action could be a choice of fertilizer recommendation. The reward might be the resulting yield or profit for that season. Unlike full RL, the contextual bandit does not consider sequential decisions over time; each decision is independent given the context. Contextual bandits have been widely used in online settings such as personalized advertising and treatment recommendations, where at each round the algorithm recommends an option based on current context and learns from the observed reward (Li et al., 2010). They are well-suited to scenarios like fertilizer recommendation if each field and season is treated as an independent decision opportunity.

The challenge in both contextual bandits and offline RL is how to learn, and evaluate, a good policy using data generated by a different policy in the past. In a typical scenario, historical agricultural data come from farmers or extension programs following their usual practices. A learned policy might suggest different actions, such as higher or lower fertilizer rates for certain conditions – but since those actions were seldom taken in the historical data, it must carefully inferred what their outcomes would have been. This is a classic off-policy learning problem. The next two sections (§2.4 and §2.5) will delve into the causal inference and off-policy evaluation techniques that address this challenge. Here, relevant work applying RL and bandit algorithms to agriculture is highlighted.

As bandit-algorithm-based adaptations to optimize crop management are becoming increasingly popular, a major issue has been the exploration-exploitation tradeoff - to find out what the best practice is, you have to try different actions (exploration), but at the same time, farmers do not want to lose money by trying something new (exploiting the current known best practice). Thus, farmers inherently experience this tradeoff — they may try an experimental fertilizer on a smaller area while continuing to grow crops with the same successful fertilizer in the rest of their fields (to limit the risk) (Gautron et al., 2024).

In addition, the researchers provide examples of how contextual bandits formally develop this concept and build upon it by including specific characteristics of each field. For example, Moothedath et al. (2023) developed a Contextual Bandit framework for a Farming Recommender System which simultaneously considers recommendations for crop varieties; fertilizer types; and irrigation amounts. The authors' formulation of the context incorporates field-specific information (location, soil type, etc.) and the action is a combined recommendation (crop choice, plus fertilizer and irrigation strategy) (Moothedath et al., 2023). The reward is defined as the Farmer's net profit. Importantly, the authors took a conservative bandit approach -- imposing constraints so that the resulting policy will never suggest an action that would result in a loss that falls below a predetermined level of performance, nor violate any agronomical constraints (Moothedath et al., 2023). This is important because, in agriculture, all recommendations must take into consideration domain knowledge.

Use of offline bandits and RL in agriculture is relatively new. Gautron et al. (2022) used offline reinforcement learning techniques, but, instead of training them in the

real-world environment, they trained them in a simulated version of the world using DSSAT (a crop simulation model). The reason they did this is to highlight the practical implications – when there is limited ability to conduct experiments in the real world, researchers can use very accurate simulations of the real world to produce large amounts of artificial data that the agent can then learn from. After the agent learns from the simulated data, the learned policies can then be tested on actual farms.

There is a growing body of research using bandit algorithms to adapt and optimize crop management. A fundamental concept is the exploration–exploitation trade-off: to learn the best practice, one must try different actions (exploration), yet in farming one also wants to minimize losses by exploiting the current best-known practice. Farmers inherently face this dilemma – they might test a new fertilizer regimen on a small plot while keeping the rest of their fields on the tried-and-true regimen, thereby limiting risk (Gautron et al., 2024).

Contextual bandits formalize this tradeoff process. Moothedath et al. (2023), for instance, present a contextual bandit framework for a farming recommender system that simultaneously considers crop variety, fertilizer type, and irrigation amount recommendations. In their framework, the context includes farm-specific data, such as location, soil properties, weather indicators, and the action is a composite recommendation made up of crop, plus fertilizer and irrigation decisions (Moothedath et al., 2023). The reward is defined as the farmer's net profit. Importantly, they formulate a conservative approach, introducing constraints so that the learned policy does not recommend actions that violate known agronomic knowledge (Moothedath et al., 2023). This ensures a level of safety and helps to establish trust. In agriculture, as in other

domains, suggestions generated from machine learning techniques must be grounded in domain expertise.

The use of offline bandit and RL algorithms in agriculture is still relatively new. Gautron et al. (2022) utilize a form of offline reinforcement learning but within simulated environments: they learned policies using the crop simulator DSSAT as a surrogate for the real world. This approach underscores a practical point – when real experimentation is limited, simulators can generate synthetic data for RL agents to train on, after which, potentially, the learned policies can be tested in field trials. However, to trust such policies, one would greatly benefit from a causal interpretation of its results, to better ensure the policies will generalize to actual farms. The following section turns to the topic of causal machine learning, which provides tools to infer cause-effect relationships from observational data, such as between fertilizer and yield.

2.4 Causal Machine Learning

Causal machine learning refers to a set of methods that combine machine learning techniques with causal inference to estimate the effects of interventions from data. Unlike predictive ML techniques, which might tell us, with high precision, the expected yield for a given fertilizer rate and field conditions, causal ML aims to answer counterfactual questions, such as "How would crop yield change if one increased the fertilizer rate?" and "What is the effect of a treatment, such as using an improved seed variety, on farmers' profits?" Answering such questions is certainly fundamental for decision support, since the primary goal is to recommend actions that cause better outcomes, not

just predict outcomes. In agronomy, observational datasets are often confounded – fields receiving more fertilizer might also have better irrigation, richer farmers, or other differences. Causal ML methods help adjust for these confounders and estimate the true impact of management decisions.

One strand of causal ML focuses on estimating heterogeneous treatment effects. Traditional average treatment effect analysis, comparing mean yield with vs. without fertilizer, for example, might hide the fact that responses vary by context. Causal forests, introduced by Wager and Athey (2018), are an example of an algorithm that extends random forests to estimate the conditional average treatment effect (CATE) for each observation (Rehill, 2025). Such methods fall under causal machine learning because they use flexible non-parametric ML models to capture how treatment effects differ across feature space (Kakimoto et al., 2022; Rehill, 2025).

Another important development is Double Machine Learning (Double ML) by Chernozhukov et al. (2016), which uses ML to estimate both the outcome model and the treatment assignment, or propensity, model, and then combines them in a way that cancels first-order errors from either model. This is an example of a doubly robust approach in causal inference (discussed more in the next section). Giannarakis et al. (2022) applied such an approach in the context of sustainable agriculture. They frame the adoption of certain sustainable practices, like crop rotation or no tillage, as a treatment and used Double ML to estimate the heterogeneous effects of these practices on soil organic carbon at the field level across Lithuania (Giannarakis et al., 2022). By leveraging climate and land-use data with ML models for both the practice-adoption process and the outcome, they aim to personalize recommendations for where sustainable

practices would be most beneficial; however, their reported results are preliminary and not statistically significant (Giannarakis et al., 2022).

Causal ML, therefore, provides a robust framework for understanding cause and effect in data-driven agronomy. It allows us to move from mere correlation (i.e., farms with higher fertilizer have higher yields) to causation (i.e., increasing fertilizer by a certain amount on a given farm would raise that farm's yield by an estimated amount). This is fundamental when deploying reinforcement learning or bandit policies on observational data: without causal corrections, an RL agent might erroneously learn that a certain action is good just because the type of farms that took that action were advantaged. By using causal ML techniques, the praxis aims to ensure that the policies learned truly reflect agronomic cause-effect relationships and will generalize to produce the intended benefits in practice. The next section deals specifically with how such learned policies are validated using off-policy evaluation methods.

2.5 Off-Policy Evaluation (OPE)

Off-policy evaluation is the task of estimating the performance of a candidate decision policy using data generated by a different policy or behavior in the past. In the context of this research, it means predicting how well a new fertilizer recommendation policy would perform, in terms of net profit, based on historical data from farmers who were following their own practices. OPE is crucial for offline reinforcement learning and bandit algorithms because it enables us to test policies before deploying them in the real

world. By only field-testing policies that are promising, potentially costly failures can be avoided.

Formally, an OPE algorithm takes as input a batch of logged data – consisting of contexts, actions taken, and rewards observed under some behavior policy – and a target policy, a new mapping from context to action. It outputs an estimate of the value of the target policy – the expected reward if that policy were used. The core difficulty is that the logs contain a biased sample of actions: they over-represent actions the logging (behavior) policy preferred and under-represent, or have no data for, actions it rarely or never took. Naively using the sample average of rewards for the target policy's actions would be biased, since the sample is not representative of the target policy's distribution (Dudik et al., 2011). For example, if the target policy often recommends higher nitrogen rates than farmers historically used, then there is little direct evidence on the outcomes of those higher rates – it cannot just be assumed they would yield similarly to the few instances in the data, which might have been tried on different fields or conditions.

To address this, two main families of OPE methods have been developed: model-based and importance sampling. Model-based approaches, often called the Direct Method (DM) in OPE, involve training a predictive model for reward given context and action and then using it to predict the rewards under the target policy for each context in the data (Dudik et al., 2011). For instance, one could train a regression to predict maize yield from features including fertilizer rate, then plug in the fertilizer rate that the target policy would choose for each field in the dataset, and average the predictions. This can yield low-variance estimates if the model is accurate, but it is prone to bias if the model is mis-specified or extrapolating beyond the support of the data (Dudik et al., 2011).

Importance sampling (IS) methods, on the other hand, re-weight the observed rewards by how "important" they are to the target policy. The simplest form is Inverse Propensity Score (IPS) weighting. If the logging policy chose action a in context x with probability $\mu\ (a \mid x)$, and the target policy would choose that same action with probability $\pi\ (a \mid x)$, then the reward is weighted by π/μ (Dudik et al., 2011). Intuitively, if the target policy likes an action more than the logging policy did, the few times that action was taken in the data should count more, by up-weighting these outcomes; conversely, if the target would rarely choose an action that the logging policy took often, those outcomes are down-weighted. When the logging policy probabilities are known, or can be estimated, IPS produces an unbiased estimator of the target policy value in theory (Dudik et al., 2011). However, IPS often has high variance, especially if the target policy diverges significantly from the logging policy. Extremely large weights can occur when an action has low probability under the logging policy but non-negligible probability under the target policy – a common situation when proposing a new strategy.

To reduce variance, a self-normalized importance sampling estimator (SNIPS), also known as weighted importance sampling (WIS), is often used, which normalizes the weights to sum to one (ZOZO Technologies, 2025). This eliminates one source of variance – the randomness in total weight – at the cost of introducing some bias. Self-normalization essentially forces the evaluated policy's weighted data to resemble a probability distribution.

Recognizing that both pure model-based and pure weighting approaches have drawbacks, hybrid estimators have been developed. The most notable is the Doubly Robust (DR) estimator (Dudik et al., 2011). In contextual bandits, the DR estimator uses

both a reward model, as in DM, and importance weighting. It evaluates the target policy's value by, roughly, taking the model's predicted reward for the target action and then adding a correction term that uses IPS to account for the difference between predicted and actual reward for the action that was taken (Dudik et al., 2011. The remarkable property of DR is that it is doubly robust: if either the reward model is accurate or the propensity (behavior) model is accurate, or both, the estimator will be unbiased (Dudik et al., 2011). In other words, even if one component is mis-specified, the other can compensate. Empirical studies by Dudík et al. (2011) showed that the doubly robust approach can achieve lower mean-squared-error than IPS or DM alone, as it trades off bias and variance effectively. DR tends to have lower variance than plain IPS when a reasonably good outcome model is available, because it uses the model's estimate as a baseline and only corrects it with weighted residuals (Dudik et al., 2011).

Off-policy evaluation is an area of active research, and many improved estimators and techniques have been proposed. Some examples include weighted doubly robust, which like SNIPS normalizes the weights in the DR estimator to stabilize it (Thomas et al., 2016), and model-based estimators for sequential (RL) scenarios such as Fitted Q-Evaluation for MDPs (Zhang et al., 2022). There are also high-confidence OPE methods that produce statistical confidence intervals or bounds on policy performance as in Thomas et al. (2015), which are valuable when making decisions from OPE estimates. Additionally, when dealing with sequential decisions as in "full" RL, importance sampling can be applied over trajectories, and more sophisticated corrections, like using marginalized importance weights or learned value functions, are needed to tackle the compounding variance and bias. The fundamental issues, however, remain similar:

correcting for distribution shift between logging and target policies and using all available information to make the most efficient estimates.

In summary, off-policy evaluation allows us to critically assess a learned policy using existing data before implementation. It is rapidly gaining popularity in real systems – for instance, companies use OPE to test new recommendation algorithms on past user data, and clinicians use it to evaluate treatment policies on retrospective patient data (Komorowski et al., 2018; Li et al., 2011). In agriculture, OPE can be used to estimate how much yield gain a new fertilizer recommendation policy might achieve, using past farm trial data, thereby avoiding unwarranted optimism or the need for extensive field trials for a new candidate policy. However, OPE itself relies on certain assumptions, notably, that the logging data has support for the target policy's actions, and its accuracy depends on the quality of models used. This motivates the use of advanced estimators like DR and its variants, which is discussed next.

2.6 Pareto-Smoothed Importance Sampling (PSIS)

The practical reliability of such estimators hinges on the tails of the weight distribution $\{w_i\}$: if a few extreme ratios dominate, the estimator can have huge variance or even fail asymptotically. Pareto-smoothed importance sampling (PSIS) addresses this by fitting a generalized Pareto distribution (GPD) to the largest weights and using the fitted tail both as a diagnostic and to stabilize the heaviest weights (Vehtari et al., 2024). The GPD's shape parameter k summarizes right tail heaviness; its estimate $\hat{k}$ is used as a convergence-rate diagnostic. In practice, PSIS replaces the top-tail weights by the

expected order statistics from the fitted GPD and then renormalizes; this preserves consistency while reducing RMSE compared with raw importance sampling or simple truncation (Vehtari et al., 2024).

2.7 Self-Normalized Doubly Robust (SNDR) and Doubly Robust (DR) Estimators

As introduced above, the doubly robust (DR) estimator is a pivotal method in off-policy evaluation and causal inference due to its error-correcting properties. In a bandit context, the DR estimator for a target policy π_1 uses two ingredients: a model $\hat{\mu}(x, a)$ that predicts reward for each action in each context, sometimes called the Q-function, and propensity scores $\pi_0(a \mid x)$ for the logging policy. For each data point with context x, action a, and reward r, DR computes the estimate as:

$$\hat{V}_{DR}(\pi_1) = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{a} \pi_1(a|x_i)\, \hat{\mu}(x_i, a) \; + \; w_i (\, r_i - \; \hat{\mu}(x_i, a_i)\,) \right) \qquad (2-1)$$

where

$$w_i = \frac{\pi_1(x_i, a_i)}{\pi_0(x_i, a_i)} \qquad (2-2)$$

are the policy's importance weights. Here $\pi_1(a|x_i)\, \hat{\mu}(x_i, a)$ refers to the model's prediction for the action the target policy would take, and the second term adjusts it by adding an importance-weighted correction. If the model $\hat{\mu}$ is perfect, the correction term has expectation zero (since $r_i - \; \hat{\mu}(x_i, a_i)$ would be zero on average), so you get the right answer. If the model is imperfect but the propensity weighting is accurate and data abundant, the correction term adjusts for the model's bias using the actual outcomes

(Dudik et al., 2011). Notably, DR yields consistent and unbiased value estimates as long as either $\hat{\mu}$ is correct or π_0 is correct – hence doubly robust. In practice, DR often performs well even if neither model is perfect, because errors from one component can cancel or be attenuated by the other (Dudik et al., 2011). By leveraging the strengths of both model-based and IPS methods, DR usually achieves lower variance than IPS, since it doesn't rely solely on raw rewards, and lower bias than a pure model, since it uses actual outcomes to correct any systematic mistakes) (Dudik et al., 2011).

Building on DR, researchers have introduced modifications to further improve stability. One such variant is the Self-Normalized Doubly Robust (SNDR) estimator. As the name suggests, SNDR applies the idea of self-normalization of importance weights within the DR formula. In a standard DR, if the logging policy sometimes took actions that the target policy favors only rarely, or vice versa, the raw importance weights w_i could be extreme. SNDR shrinks the importance weights by dividing by their sum, similarly to how SNIPS normalizes weights in pure importance sampling (Saito et al., 2021). The motivation is to reduce variance: no single data point can overly dominate the estimate because the weights are scaled to sum to 1. This often yields a more stable estimate of policy value, particularly in cases of heavy-tailed weight distributions or small sample sizes; the trade-off is that SNDR is no longer strictly unbiased, even if one of the models is correct – the normalization introduces a slight bias (Swaminathan et al., 2015). Mathematically, if w_i (as defined in Equation 2-2 above) is the importance weight for observation i, SNDR would use normalized weights in the correction term (Saito et al., 2021), giving:

$$\hat{V}_{SNDR}(\pi_1) = \frac{1}{n}\sum_{i=1}^{n}\sum_{a}\pi_1(a\,|x_i)\,\hat{\mu}(x_i,a) \;+\frac{\sum_{i=1}^{n} w_i\;(r_i - \hat{\mu}(x_i,a_i))}{\sum_{i=1}^{n} w_i} \qquad (2-3)$$

Beyond SNDR, two other widely used methods are DR with optimistic shrinkage (DRos) and Switch-DR, which borrow ideas from both machine learning and causal statistics to further control bias-variance trade-offs (Su et al., 2020; Wang et al., 2025). Switch-DR, for instance, uses the direct method when the importance weight is above a threshold, and switches to DR otherwise (Wang et al., 2025).

In summary, Doubly Robust (DR) estimation provides a strong foundation for off-policy evaluation by requiring only one of two estimators, model or propensity, to be consistent. Its self-normalized variant SNDR improves practical stability by taming extreme weights at the cost of a small bias. Both are highly relevant to this dissertation's methodology. In Chapter 3, doubly robust estimators are utilized to evaluate and train policies, ensuring that the fertilizer recommendation policy can be assessed with greater confidence using existing data. These estimators will help us make the most of the available agronomic data, which is often biased by farmers' current practices, by combining predictive modeling with principled re-weighting. The literature reviewed in this chapter thus provides the conceptual underpinnings – from agronomic decision support and crop modeling, through reinforcement learning and causal inference, to advanced evaluation techniques – that collectively inform the development of this study's approach. The next chapter will build on these insights to propose a methodology that addresses the identified gaps: using one-step offline RL (contextual bandit) methods enhanced with causal ML and OPE techniques to deliver a robust, data-driven fertilizer decision support solution.

Chapter 3—Methodology

3.1 Introduction

This chapter details the end-to-end methodology for developing the contextual bandit fertilizer recommendation system. The methodology begins with data cleaning, and it then crafts features representing the decision context (soil, topography, and pre-plant climate), without leaking future information. The fertilizer action space is then discretized into a finite set of fertilizer strategies using adaptive binning techniques to ensure actions are supported by the data and agronomically interpretable. A surrogate crop profit model is constructed using machine learning techniques (a stacked ensemble of XGBoost, LightGBM, and CatBoost) to predict profit from context and fertilizer inputs. With this surrogate model, a contextual bandit optimization framework is applied to derive a fertilizer policy and ensure conservative overlap criteria are met to avoid unsafe extrapolation. Rigorous off-policy evaluation with doubly robust and self-normalized doubly robust estimators and bootstrap confidence analyses is then conducted to estimate the economic value of the policy. Finally, the optimized policy is translated into a decision support tool, in which the user can input their given farm conditions, and the tool recommends optimized fertilizer amounts, complete with safeguard mechanisms. Figure 3-1 provides a concise flowchart of this study's methodology.

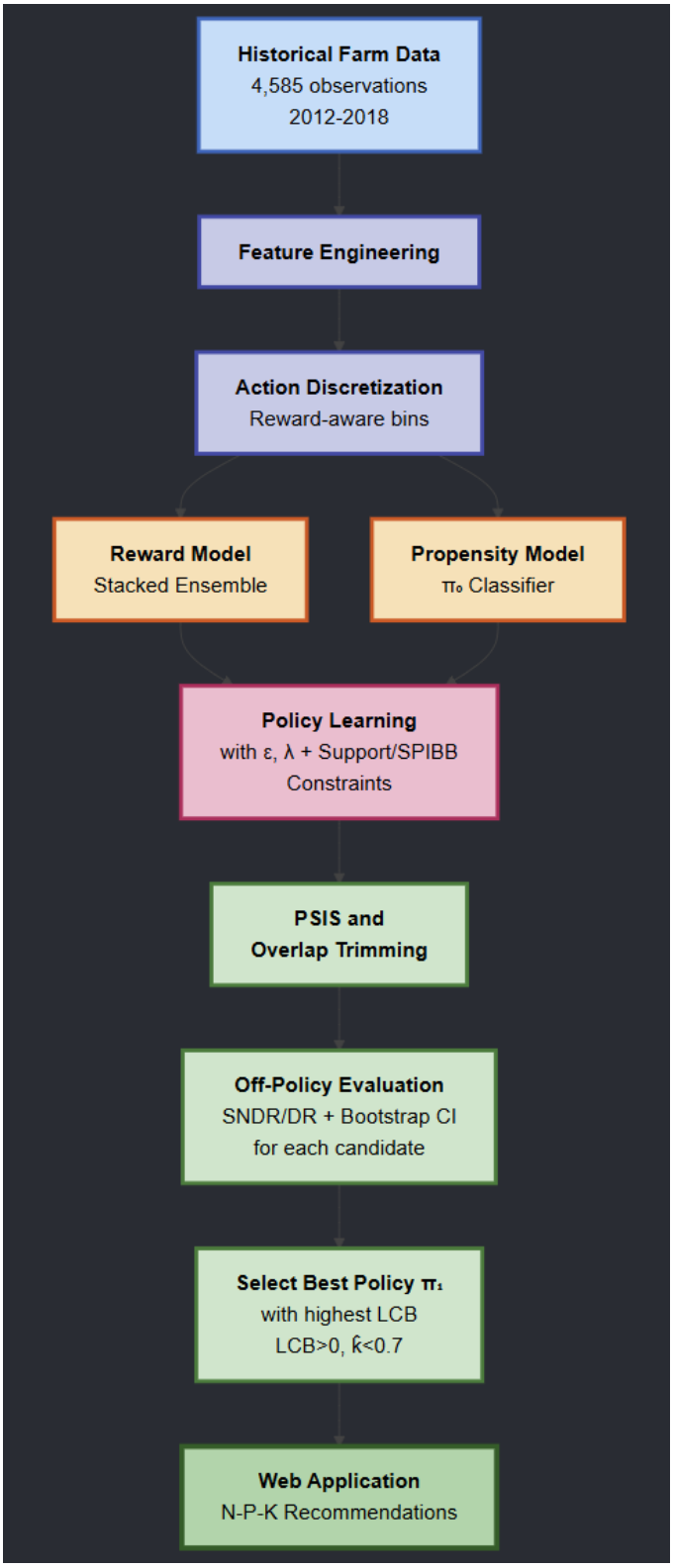

Figure 3-1. Flowchart of methodology

3.2 Data Collection and Preprocessing

Data Source and Description. This study utilizes a large multi-year dataset that was collected by the International Maize and Wheat Improvement Center (CIMMYT) over seven years (2012–2018) from on-farm maize trials in Chiapas, Mexico, (Trevisan et al., 2022). The dataset contains 4,585 field-level observations, each representing a particular farmer's field during a specific season. For each observation, the dataset records maize yield and a rich array of contextual variables encompassing management, soil, topography, and weather conditions. Table 3-1 below summarizes the variables included in the dataset:

Table 3-1. Variables included in the Chiapas dataset (Trevisan et al., 2022)

Variable	Description	Unit
Field_ID	Unique identifier	
Lat	Latitude coordinates	Degrees
Long	Longitude coordinates	Degrees
Elev	Elevation	m
Mun_Yield	Average maize yield in a municipality	Mg ha^{-1}
System	Specification of hybrid or landrace-based cropping system	
Yield	Field maize yield	Mg ha^{-1}
Cultivar	Genotype information	
Tillage	Tillage practices including Conventional, Reduced, and No-Till	
Planting	Day of the year when maizes was planted	DOY
Nitrogen	Nitrogen added as fertilizer	kg N ha^{-1}
Phosphorus	Phosphorus added as fertilizer	kg P_2O_5 ha^{-1}
Potassium	Potassium added as fertilizer	kg K_2O ha^{-1}
Slope	Field slope	%
Clay	Soil clay content	%
CEC	Cation Exchange Capacity	cmolc dm^{-3}
SOM	Soil Organic Matter	**%**
pH	Soil pH	
prcp(V1-V30)	Precipitation per day over 10 days from V1 to V30	mm day^{-1}
srad(V1-V30)	Solar radiation per day over 10 days from V1 to V30	MJ m^{-2} day^{-1}
tmax(V1-V30)	Maximum temperature over 10 days from V1 to V30	°C
tmin(V1-V30)	Minimum temperature over 10 days from V1 to V30	°C
Vp (V1-V30).	Vapor pressure over 10 days from V1 to V30	Pa

Maize yields range from as low as 0.1 up to 10.0 Mg/ha, with 75% of observations below 5.0 Mg/ha, demonstrating substantial yield variability that is common in tropical smallholder systems (Trevisan et al., 2022). The data includes an extensive array of farmer management practices and environments; for example, elevation varies from near sea level to about 3000 m across fields (with the median around 700 m), and planting dates span nearly the entire rainy season (with a standard deviation of 20 days, and some sowing dates up to 5 months apart) (Trevisan et al., 2022). These variations highlight the need to account for site-specific context in the recommendation system.

Management and Input Variables. Each observation includes the farmer's fertilizer application rates of nitrogen (N), phosphorus (P_2O_5), and potassium (K_2O) in kg/ha, as well as categorical variables describing the cropping system and tillage practice. The *System* variable indicates whether the field utilizes a semi-commercial hybrid or a smaller-scale landrace cropping system. This distinction is important because hybrid systems, concentrated in central-west Chiapas, generally achieved higher yields (often 4.2–5.3 Mg/ha) whereas landrace systems in northern/eastern areas yielded lower (1.6–3.0 Mg/ha). The dataset confirms that hybrid-system farmers tended to have better resources and higher productivity, whereas landrace farmers were more subsistence-oriented. *Tillage* is recorded with three levels: conventional plowing (the most common, 45% of cases), no-till (36%), and reduced tillage (19%). These management factors are considered as context features affecting yield response to inputs.

Crucially, input use in the dataset is highly variable. Nitrogen fertilizer application ranges widely – the average N rate was 110 kg N/ha, but some farmers applied up to 330 kg N/ha (three times the mean). A nontrivial fraction of fields

(including some relying on manure or unreported inputs) had effectively zero synthetic N applied. Phosphorus and potassium use were even more skewed: over 25% of farmers used no P fertilizer at all, and more than 50% applied no K fertilizer. Those who did use P and K applied moderate amounts (as seen in Table 3-2 below), but zero-usage is a dominant category for these nutrients. This sparsity in P and K usage will inform how these inputs are discretized and modeled. Additionally, the data include *Cultivar* (genotype identifiers); however, with 250 unique cultivars (most appearing only a few times), cultivar data are extremely granular. Cultivar information is treated, therefore, at a higher level (hybrid vs. landrace via the System variable) to avoid overfitting to rare variety effects.

Table 3-2. Descriptive statistics of the Chiapas dataset (Trevisan et al., 2022)

Variable	Mean	SD*	P0	P25	P50	P75	P100	Hist
Yield (Mg ha^{-1})	6	1.9	0.1	1.9	6	5.0	9.8	
Elevation (m)	884	443	7	592	712	1079	2849	
Slope (%)	6.0	6.8	0	1.3	3.0	8.7	61.2	
Clay (%)	30	10	5	23	28	37	57	
CEC (cmolc dm^{-3})	22.7	7.4	4.3	16.2	21.1	27.6	50.8	
SOM (%)	1.6	0.9	0.3	1.0	1.2	2.0	4.0	
pH	6.8	0.7	4.9	6.6	6.8	7.3	8.3	
Planting (DOY)	165	22	91	153	167	180	242	
Nitrogen (kg N ha^{-1})	109	64	0	64	110	156	349	
Phosphorus (kg P_2O_5 ha^{-1})	23	26	0	0	23	46	143	
Potassium (kg K_2O ha^{-1})	9	17	0	0	0	12	100	
Precipitation (mm day^{-1})	4.05	4.86	0	0	2.1	6.8	54.9	
Solar Radiation (MJ m^{-2} day^{-1})	17.7	2.8	6.8	15.6	17.4	19.5	26.8	
Maximum Temperature (C^{o})	29.8	2.5	16.6	28.4	30.2	31.5	39.2	
Minimum Temperature (C^{o})	17.4	1	-2.1	15.5	18.0	19.7	36.2	
Vapor Pressure (Pa)	1761	628	252	1336	1900	2240	6452	

*SD: standard deviation; P0 – P100: data distribution percentiles; CEC: cation exchange capacity; SOM: soil organic matter; DOY: day of the year.

Soil and Topographic Features. Each field's soil properties are recorded, including %Clay content, soil pH, Cation Exchange Capacity (CEC in cmol_c/dm³), soil organic matter (SOM %), and slope (%). These features represent the edaphic context influencing nutrient availability and crop growth. Elevation and geographic coordinates (latitude, longitude) are provided for each field, enabling topographic context. Elevation is included as a continuous feature; it captures temperature and orographic rainfall gradients and indeed shows a negative correlation with yield in this region (Bhat et al., 2024); higher elevation fields tend to yield less, partly due to cooler climate and more resource-constrained farmers. Longitude and latitude are used both as input features and to form ≈1 km geo-tiles that group nearby fields during model training. The municipal-average yield variable (*Mun_Yield*) is explicitly excluded from the feature set to avoid outcome leakage.

Weather Features. A rich set of weather variables was collected from external sources (Daymet and national databases) for each field's growing season. Weather data are provided as daily values aggregated into 10-day intervals across the season: for each decadal period (V1 to V30) the dataset includes total precipitation (mm/10days), average solar radiation (MJ/m²/day), and average daily minimum and maximum temperatures (°C), and vapor pressure (Pa). In total, this yields up to 30×5 = 150 weather features per field, representing the time series of climate during the crop cycle. The average length of a season is ≈175 days (≈18 ten-day periods from planting to harvest on average), but because planting dates differ and some fields may have longer crop durations or delayed harvest, the dataset includes a fixed 300-day window (30 decadal periods) to cover all cases. It's important to note that many of these weather features occur after fertilizer

application decisions, especially if fertilizer is applied around planting or early growth. In constructing the decision-support context, the inclusion of *ex-post* weather data is avoided that would not be available at the time a farmer chooses a fertilizer strategy (see *Leakage Avoidance* below). Instead, only weather up to the decision point (i.e., the first 6 decadal periods) is included as context for recommendations.

Data Cleaning and Preparation. Standard preprocessing steps to prepare the data for modeling are performed. First, the dataset was checked for completeness and consistency; however, the dataset was largely complete due to the systematic data collection process by Trevisan et al. (2022). The dataset was found to be generally clean; extreme yield values aligned with known high-performing cases, and extremely high N rates were rare and typically documented.

Leakage Avoidance: A critical aspect of preprocessing was ensuring that the modeling features represent information available at or before the decision time for fertilizer application. The objective here is to avoid data leakage, where the model could otherwise cheat by using future knowledge (such as the weather after planting) to make recommendations. It is assumed that the fertilizer recommendation is made around planting time; this aligns with typical practice where basal fertilizer is applied at planting and possibly adjusted with a top-dress a few weeks later (since the dataset does not contain top-dress fertilizer values, a one-time application of fertilizer is assumed when formulating the problem). Therefore, features that are only known after planting, such as mid-season rainfall or temperatures, cannot be used for the ex-ante decision. To enforce this, all weather variables beyond the planting date are excluded from the policy's context

features. Additionally, the *Mun_Yield* (municipality average yield) variable is also explicitly omitted from the feature set.

Selecting only ex-ante features and encoding variables appropriately establish the foundation for feature engineering, modeling, and policy learning. The following sections detail how covariates from these raw features are constructed, how the profit outcome is modeled, and ultimately how a target fertilizer recommendation policy is derived.

3.3 Feature Engineering and Covariate Construction

With the data prepared, features are then engineered to best capture each field's context at the time of the fertilizer decision. The goal of feature engineering is to represent the state of the farming system in a form that is both predictive of yield and available prior to fertilizer application, drawing on agronomic domain knowledge and data analysis techniques.

Temporal Encoding of Planting Date. Planting day-of-year (DOY) is an important variable in rainfed systems because it captures the timing of the cropping season relative to rainfall patterns. Rather than use *DOY* as is, it is encoded in a way that respects its cyclical nature; the last day of the calendar year wraps to the first day of the calendar year (Pewsey et al., 2021). Two features are thus engineered, $\sin\left(2\pi \cdot \frac{DOY}{365}\right)$ and $\cos\left(2\pi \cdot \frac{DOY}{365}\right)$, which effectively map the planting date onto a unit circle. This cyclical encoding ensures that, for example, a planting date of late December and another of early

January are recognized as temporally proximate in feature space, rather than maximally distant.

Of note, to capture yearly trends, *Year* is kept as a covariate in the feature matrix used for model training, rather than dropping it. Generalization across years is addressed via forward-year cross-validation, which is detailed in later sections.

Weather Aggregation and Soil Features. Weather features are restricted to pre-planting windows only and explicitly exclude any post-plant information to avoid data leakage. Concretely, weather variables are aggregated over V1–V6 (60 days before planting in total, each V representing a decad of 10 days). For precipitation, *sum*, *mean*, *std*, *min*, *max* are computed, and for solar radiation, maximum and minimum temperature, and vapor pressure *mean*, *std*, *min*, *max* are computed. A simple proxy for dry spells, the count of dry decads, defined as <1 mm/day across V1–V6, is also added.

The soil variables *pH*, *SOM*, *clay*, *CEC*, and *slope* are used essentially as provided, with no further feature engineering done.

Categorical Management Feature. The *System* and *Tillage* variables are one-hot encoded. Since it is likely that the optimal fertilizer application rate will depend on the type of system, including the *System* as a feature allows the model to suggest different actions based on whether the farmer has hybrid or landrace seeds. In addition, since it is likely that fertilizer application efficiency might vary depending on the type of tillage used (i.e., conventional vs. no-till vs. reduced), *Tillage* is included as a feature to account for these interactions. For instance, farmers who use no-till systems may have more surface residue and possibly different nitrogen dynamics than those who use conventional

or reduced tillage systems, thus potentially affecting their optimal nitrogen application levels.

Cultivar is not included as a feature due to the high number of unique maize varieties (totaling 250) with some being very rare, which would likely result in overfitting when attempting to one-hot encode them. Not including of *Cultivar* as a feature avoids the classic "curse of dimensionality" associated with sparse categorical data.

Action Representation (Fertilizer Inputs): In the Chiapas dataset, N, P, and K application rates are continuous variables. In preprocessing, N, P, and K are excluded from the covariate matrix X and store them separately as a 3-column continuous actions array used by the reward model. During training, N, P, and K are kept continuous for model fitting. Separately, adaptive bins for N, P, and K are learned from the training data and use the binned actions for diagnostics such as overlap and coverage. The binning is purely data-driven and may differ by cross validation fold.

To conclude, the feature engineering performed provides an array of covariates for describing fields relative to their topography, soil fertility, the timing of planting, pre-plant climate, and the associated farm management regimes. It has been demonstrated by prior work, as stated above, that it is important to consider all of these factors together as crop yield is influenced by many different types of interactions that occur among farm management decisions and the environment (Bhat et al., 2024). Features are limited to information available at or before the time of the fertilizer application decision point; all post-plant data and proxy data for the target variable are thus excluded. Overly-specific variables that hinder generalization are also avoided. The engineered features

consequently give the surrogate profit model and the contextual bandit policy meaningful environmental and farm management context.

3.4 Action Space Discretization - Adaptive Binning of Fertilizer Rates

An important consideration in formulating the fertilizer recommendation problem as a contextual bandit is the continuous nature of the action variables. Classic contextual bandit algorithms assume a finite set of discrete actions. To apply these methods and to ensure robust policy learning from limited data, the fertilizer rates are discretized into a finite action space (Dougherty et al., 1995). The proposed discretization approach is adaptive and reward-aware, meaning the bin boundaries were chosen based on the yield response data rather than arbitrary intervals, in order to capture meaningful differences in outcomes.

Need for Discretization. Continuous action spaces in offline reinforcement learning or bandit settings pose serious difficulties, particularly with limited data. Estimating propensities or rewards for an uncountable number of actions is infeasible, and many actions might never be tried in the data. By discretizing, the action space is restricted to a manageable set of options that have sufficient support in the dataset. This also addresses the support deficiency problem – by ensuring each possible action bin has been taken by some farmers in the data, recommending actions that are completely out-of-sample is avoided (Sachdeva et al., 2020). In practice, discretization means the system will recommend fertilizer rates in a few broad categories (i.e., “no N”, “low N”, “medium N”, “high N”) rather than any arbitrary number, which is also easier to communicate to farmers.

For each nutrient $d \in \{N, P, K\}$, a shallow decision tree regressor is fitted with the nutrient rate as the sole feature and yield as the target on training years only. The tree's split thresholds define candidate breakpoints. The number of leaves are capped per nutrient – N up to 6 bins, P up to 3, K up to 2 – and require a minimum number of samples per bin to prevent tiny, unstable bins. If the extracted thresholds don't exactly meet the target count, edges are backfilled using quantiles, and, if needed, uniform spacing, between the min and max observed rates. Binning is then performed with numpy.digitize. This procedure yields data-driven, reward-aware bins. This design accommodates the empirical distributions in the dataset, while keeping the number of actions tractable.

3.5 Reward Modeling – Surrogate Prediction Model

To enable data-driven decision optimization, a reliable model of the reward function is needed – in this case, net profit, subtracting fertilizer costs from maize revenue – as a function of context and action. This study constructs a surrogate model that predicts profit Y given the context features X (soil, weather, management practices, etc.) and a candidate fertilizer action A (discretized N-P-K combination). This surrogate serves multiple purposes: (1) it provides insight into how different factors and fertilizer rates drive profits, (2) it allows counterfactual prediction of yields for actions not taken in a particular field (which is crucial for evaluating and optimizing policies offline), and (3) it can be directly used to recommend the estimated best action for a new context (the argmax of the predicted profit).

Prices. To compute net profit, the surrogate profit model assumes fixed prices (N=16, P_2O_5=12, K_2O=8 MXN/kg, maize 3.5 MXN/kg) across all years, due to difficulty obtaining precise price data during this period. These prices were informed by *inegi.org.mx*. In reality, fertilizer and grain prices fluctuate; this is a known limitation in this study as stated in Chapter 1.

Model Choice. The praxis employs a stacked ensemble regression model for profit. Ensemble models combine multiple base learners to improve predictive performance, leveraging their complementary strengths. The ensemble $\hat{\mu}(x, a)$ is comprised a mix of three first-stage learners:

1) XGBoost (gradient boosting trees),
2) LightGBM (gradient boosting with leaf-wise growth),
3) CatBoost (ordered boosting),

followed by a RidgeCV meta-learner that linearly combines out-of-fold predictions from the base models. Before stacking, mutual information (MI) -based feature selection is performed, selecting the top 40 features after robust scaling, and then concatenating the selected features with the candidate action (N, P, K) so that the models learn interactions between context and dosage directly (Breiman, 1996; Hoerl et al., 1970; Vergara et al., 2014).

The choice of XGBoost, LightGBM, and CatBoost as base learners in the stacked surrogate model reflects both the structure of the Chiapas dataset and current best practice for tabular prediction problems. The reward model must capture rich, nonlinear interactions between soil and weather conditions, management practices, and discretized

$N–P_2O_5–K_2O$ rates, under substantial noise and modest sample size. This design allows the surrogate to exploit complementary inductive biases across the base learners while explicitly regularizing the final combination.

Specifically, XGBoost contributes robust regularization and a depth-wise growth strategy that helps stabilize predictions against the high variance inherent in the data (Chen et al., 2016). LightGBM complements this via its leaf-wise growth algorithm, which efficiently captures complex, fine-grained interactions between variables that depth-limited approaches might overlook (Ke et al., 2017). Finally, CatBoost employs symmetric decision trees, introducing a distinct structural bias that further mitigates overfitting and enhances generalization reliability on this finite dataset (Prokhorenkova et al., 2018).

Using a stacked ensemble of boosted-tree models is consistent with broader evidence that gradient-boosted decision trees remain among the strongest performers for tabular data across many domains, often rivaling or exceeding deep neural networks while retaining relatively straightforward interpretability and deployment (Shwartz-Ziv et al., 2022). This combination therefore offers a pragmatic balance between predictive accuracy, robustness under limited agronomic data, and operational feasibility, making it a suitable backbone for the contextual bandit policy and subsequent off-policy evaluation.

Training and Validation Strategy. The surrogate is evaluated in a strictly forward-by-year manner. For each validation year y, the training set is all rows with $Year < y$ and the test set is the rows with $Year = y$. Before any fitting in that year, adaptive fertilizer bins are learned only on the training years to define the joint

$N \times P \times K$ grid and to compute per-cell support counts—this preserves train/test separation and matches the support used later in OPE.

Within the held-out year y, site-based folds are built by shuffling unique geo site keys (rounded lat/lon) and splitting them into k buckets. For each test fold, a FIT set is created that is either (i) the past-years TRAIN only (for strictly prospective claims), or (ii) TRAIN ∪ (other TEST folds); the latter being appropriate for retrospective, offline policy evaluation, which is the method used in this praxis. The stacked reward model μ is then fitted on the FIT set with recency weighting to emphasize more recent seasons, and score the logged actions on the held-out fold. Fold predictions are aggregated to produce year-level R^2 and RMSE. This procedure is both temporally aware, with no future leakage, and spatially aware,

Performance Metrics. The surrogate model's predictive performance is evaluated using R^2 (coefficient of determination) and Root Mean Square Error (RMSE) on validation sets. R^2 measures the fraction of variance in profit explained by the model, given by the following equation:

$$R^2 = 1 - \frac{\Sigma_i (r_i - \hat{\mu}(x_i, a_i))^2}{\Sigma_i (r_i - \bar{r})^2} \qquad (3-1)$$

where r_i are the observed profits, $\bar{r}$ is the average profit in the dataset, and $\hat{\mu}(x_i, a_i)$ are the profit values estimated by the surrogate.

In the context of this research, an R^2 in the range of 0.4–0.6 is considered reasonably good given the high variability and unobserved factors in on-farm crop profits. RMSE, given by Equation 3-2 below, provides an absolute error scale; for

example, an RMSE of 3000 MXN/ha means the typical prediction error is ±3000 MXN/ha, which is about 29% of the average net profit (10,457 MXN/ha).

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\Sigma_{\mathrm{i}}\left(r_{\mathrm{i}} - \hat{\mu}(x_i, a_i)\right)^2} \qquad (3-2)$$

Auxiliary yield surrogate for off-policy evaluation. While the policy is learned to maximize profit, yield uplifts are also reported to separate agronomic response from price effects. To do so, a second stacked ensemble $\hat{v}$, identical to the profit surrogate but trained on yield as the target, is fitted using the same features, adaptive N–P–K bins, site-based cross-fitting, and recency weighting. This is used later during evaluation to compute yield gains for the profit-optimized policy.

Once trained, the surrogate model $\hat{\mu}(x, a)$ can predict profit for any given context x and discrete action a. This is the cornerstone for policy learning: $\hat{\mu}$ can be used to simulate counterfactual outcomes. It's important to note that $\hat{\mu}$ is an approximation – it may not perfectly reflect reality, especially in extrapolated regions. This is mitigated by confining to in-support actions. Thus, the surrogate profit model is a robust ensemble that captures the relationship between context, fertilizer, and yield. It forms the backbone of the recommendation system – first by guiding the policy where to search for profit gains and ranking actions for each x, and second by providing the expected reward estimates needed for off-policy evaluation. Having established $\hat{\mu}$, the next step is learning the actual fertilizer recommendation policy (the contextual bandit agent) that uses this model, and the original data, to optimize decisions.

3.6 Policy Learning – Contextual Bandit Optimization

With a surrogate reward model in hand and a discrete action space defined, the fertilizer recommendation task is framed as a contextual bandit problem. In reinforcement learning, at each decision point an agent observes some context $X = x$ (the field-specific features) and must choose an action $A = a$ (one of the discrete fertilizer rate combinations) to maximize the expected reward (maize revenue minus fertilizer costs). After choosing, if learning online, the agent would receive the reward of that action and update its policy. In this study, however, learning is offline: a policy is learned from the batch of historical data, rather than by interacting in real-time, and as the problem is framed as a contextual bandit, there is but one decision point, namely the time before planting when fertilizer is applied.

The Propensity Model. The first step is to learn the logging policy, i.e., the policy that generated the data. In this dataset, the logging policy (this will be used interchangeably with baseline policy and behavior policy) is not a single known policy but rather a mixture of farmer decisions and possibly extension advice. It's reasonable to assume that certain contexts led to certain actions more often (i.e., hybrid system fields likely had higher fertilizer rates applied; landrace fields often had none). These tendencies need to be estimated to later weigh the data for the new policy.

The logging policy $\pi_0(a \mid x)$ is modeled as a joint classifier over the full (N, P, K) action grid, treating each discretized triplet as a single class. For each evaluation year y, the logging-policy model $\pi_0(a \mid x)$ is refit on that year's training data and use site-based cross-fitting to obtain predictions on the test year, in a similar fashion to surrogate model training. Concretely, adaptive fertilizer bins (N, P, K) are learned

from training years only **($< y$)**, then bin actions for both train and test. π_0 is modeled as a multiclass classifier over the joint (N, P, K) grid using XGBoost. Training is performed in a compact label space containing only the joint cells observed in the fit data; at inference probabilities are expanded back to the full grid and assign exactly zero probability to never-observed cells. Recency weights are applied to emphasize recent seasons. K-fold cross-fitting is performed on the test year, sites defined by rounded latitude/longitude: each fold is fit on TRAIN ∪ (other TEST folds) and predicted on the held-out fold (as in surrogate model training); final test-year quantities aggregate predictions across folds. Isotonic probability calibration is applied but only when supported by the data: 3-fold calibration is used if the minimum per-class count among observed joint cells is ≥ 3, 2-fold if ≥ 2, and otherwise the classifier is left uncalibrated. For each context x, the model returns a probability vector over all joint actions.

For each year, expected calibration error (ECE) is reported with reliability curves, the 1st and 5th percentiles of π_0 to characterize the left tail that drives importance weights, histograms of π_0 per evaluation year, and an action-coverage summary, consisting of the number of observed joint cells and the number of cells with at least 30 counts, the specified support threshold.

Contextual Bandit Formulation: Now that the baseline policy is formulated, the next step is to formulate the contextual bandit policy. Formally, the policy is a mapping $\pi: X \rightarrow \Delta(A)$ that maps contexts to a distribution over the action set.

The challenge is that the dataset contains only samples of certain (x, a) pairs from the logging policy (the historical behavior of the farmers). Unlike supervised learning, it

cannot directly be observed what yield would have been if a different action was taken on the same context. This is why the surrogate model and off-policy evaluation are needed. This approach to policy learning blends two strategies: a model-based optimization using the surrogate to propose a candidate policy, and a conservative constraint using off-policy evaluation insights to adjust that policy for safety and support.

Initial Policy Derivation. As a starting point, a deterministic policy π_{greedy} id derived that selects, for each context, the fertilizer action that maximizes the predicted profit from the surrogate model. In other words,

$$\pi_{greedy}(x) = \arg\max_{a \in A} \hat{\mu}(x, a) \qquad (3-3)$$

where $\hat{\mu}$ is the ensemble reward model. This can be done by enumerating all feasible actions *a* (the 36 bins combinations) and computing the predicted yield for each given the context *x*, then picking the top-yielding one. Because the action space is small, this brute-force argmax is trivial to compute. The resulting policy is essentially the best fertilizer strategy according to the learned yield response surface. For instance, π_{greedy} might learn that for a hybrid maize field with high elevation, low SOM, and early planting, the best action is "High N, medium P, high K", whereas for a landrace at low elevation with good soil, the best is "Medium N, low P, no K", etc., depending on what maximizes predicted yield. This approach leverages the full model information; however, it can be prone to exploiting model errors: if $\hat{\mu}(x, a)$ overestimates yield for a rare action *a* in context *x*, the greedy policy might choose *a* even if in reality that action could perform poorly. This is a form of overfitting known as the optimism bias in model-based policy optimization.

Conservative Policy Constraints: To counteract potential over-optimism and to keep recommendations within known-safe bounds, the learned policy is constrained so it stays close to historical practice. Concretely, three mechanisms are employed that operate strictly on train-supported actions: (i) a support constraint, (ii) on-support ε-greedy exploration, and (iii) baseline mixing.

- *Supported-Only Actions:* The policy is restricted to recommending actions that lie in the support of the logging data for similar contexts: A binary support mask is defined over the joint N–P–K grid from the training data; a cell is "supported" if it has at least 30 logged samples, otherwise it is masked out. The target policy π_1 is then constructed per row to assign probability only to cells where both the support mask is true and the estimated logging propensity π_0 is positive. This support-aware strategy prevents extrapolation into never-logged combinations and stabilizes importance weighting.
- *Safe Policy Improvement with Baseline Bootstrapping (SPIBB):* As an additional guardrail, a safe action constraint inspired by SPIBB is imposed. SPIBB theory (Laroche et al., 2019) dictates that the learned policy should defer to the baseline in any state-action pair that is insufficiently supported by data. In the context of this study, this translates to: if a certain fertilizer bin has very few samples in the data, the policy is not allowed to deviate from the baseline behavior in that bin. This is enforced by identifying all joint bins with sample count less than a minimum – which is defined as 60 samples – and for those bins, if the argmax policy chooses them for some context, it instead defaults to the baseline action for that context.

- *Incorporating Exploration (ε-Greedy on Support)*: Controlled exploration is injected while staying strictly within train-supported actions. For each context x, the best supported action is taken under the profit objective and a small, uniform exploration mass is injected over supported cells:

$$\pi_{\varepsilon}(\cdot \mid x) = (1-\varepsilon)\,\delta_{g(x)} + \varepsilon\, U_{support}, \qquad (3-4)$$

where $U_{support}$ is the uniform distribution over supported actions. If $S(x)$ is the set of actions marked supported at x, then

$$U_{support}(a \mid x) = \frac{1}{|S(x)|} \qquad (3-5)$$

for $a \in S(x)$, and 0 otherwise. Here $\delta_{g(x)}$ is a point-mass Dirac distribution that puts all probability on the greedy supported action for context x. If

$$g(x) = \arg\max_{a \in S(x)} \hat{\mu}(x,a), \qquad (3-6)$$

then $\delta_{g(x)}(a) = 1$ if $a = g(x)$, else 0.

- *Baseline Mixing:* This exploratory distribution is then mixed with the masked baseline policy:

$$\pi_1(\cdot \mid x) = \lambda\, \pi_0^{\text{supp}}(\cdot \mid x) + (1-\lambda)\, \pi_{\varepsilon}(\cdot \mid x), \qquad (3-7)$$

where π_0^{supp} is π_0 renormalized on supported cells and $\lambda \in [0,1]$ controls conservatism. Larger λ keeps the policy closer to historically practiced actions; smaller λ leans into the surrogate-driven recommendations. The choice of λ is tuned via off-policy evaluation, subject to overlap diagnostics.

The final learned policy π_1 can be described as a conservative contextual bandit policy: it mostly follows the action that maximizes predicted profit, but it's tempered by domain constraints. By being careful in this way, this praxis aims to deliver recommendations that are not only high-yielding in theory but also credible and low-risk in practice.

Overlap Trimming. A major concern in OPE is lack of overlap – if the new policy places weight on actions that the behavior policy rarely or never took in certain contexts, the importance weights will be large or undefined, making the estimates unreliable. To address this, a form of trimming (Crump et al. 2009) is applied to define an overlap subset of the data where there is sufficient support. Several criteria are implemented to trim out high-risk samples:

- *Minimum behavior probability:* Any data point where $\pi_0(a_i|x_i)$, the propensity of the logged action, is below a small threshold (namely 0.01) is dropped. These are cases where the farmer's action was extremely unusual given context, indicating that context region might be under-sampled.

- *Maximum importance ratio:* Points are dropped where the raw importance weight $w_i = \frac{\pi_1(x_i,a_i)}{\pi_0(x_i,a_i)}$ exceeds a cap of 10.0. This directly removes instances that would overly skew the estimator (similar to truncating weights, which can reduce variance).

- *Minimum action count:* The logged action's bin is required to have at least a certain count in the dataset; 30 is used as the cutoff. This is different than the support mask defined above; the support mask constrains where π_1 may place

probability, not what actions were logged. If either the action actually taken or the action being recommended in that context's bin is too rare, with < 30 samples, that sample is considered outside reliable support and is excluded.

Only data points that pass all these filters are kept in the overlap subset. The retained fraction is reported as a diagnostic. This approach is analogous to the suggestion by Crump et al. (2009) to discard units with extreme propensity scores to stabilize treatment effect estimation. By trimming out the worst-off-support cases, some bias is sacrificed – not evaluating policy in poor-overlap regions – in exchange for much lower variance in the regions that are evaluated. In essence, a conservative estimate of policy performance in the well-supported regions is obtained, which is appropriate for decision-making – it is preferred to confidently report gains on a reduced subset than to overestimate them from unreliable extrapolation on a larger subset of the data.

Pareto Smoothed Importance Sampling (PSIS). Before conducting off-policy evaluation (OPE), sort the raw importance ratios, and fit a generalized Pareto distribution (GPD) to the top tail to obtain its shape parameter $\hat{k}$. The value of $\hat{k}$ directly reflects the effective number of finite moments of the raw weight distribution and hence the pre-asymptotic behavior of importance sampling. Empirically if $\hat{k} < 0.7$, the weighted estimate is moderately reliable; but if $\hat{k} \geq 0.7$, the variance of the weights is so large that the estimate becomes unreliable (and if $\hat{k} \geq 1$, the weights have infinite variance, meaning the estimate of mean reward may not even converge) (Vehtari et al., 2024). $\hat{k}$ is therefore strictly required to be < 0.7 for the evaluation of the policy. If $\hat{k} > 0.7$, the evaluation is considered unreliable and the policy would be rejected.

After confirming the reliability of $\hat{k}$, Pareto Smoothed Importance Sampling (PSIS), a weight-stabilization scheme, is performed, which replaces the largest ratios by expected GPD order statistics (a smooth Winsorization), and renormalizes to form self-normalized weights. This retains the desirable properties of self-normalized IS (consistency; finite variance under mild conditions) and empirically reduces mean-squared error relative to raw importance sampling or simple weight clipping (Vehtari et al., 2024).

The next section will detail how this policy is evaluated offline to estimate its performance and ensure these constraints were effective.

3.7 Off-Policy Evaluation

After learning a candidate policy, a crucial step is to evaluate how this policy would perform if implemented, using only the existing offline data. This is the off-policy evaluation (OPE) problem: estimating the expected reward of a new policy π_{new} given data collected under a different policy. In this study, expected farmer profit that the recommendation policy π_1 would achieve is estimated, and it is compared to the baseline policy π_0 – farmers' historical behavior that is reflected in the data.

Confidence Intervals (CIs) via Clustered Bootstrap. For each evaluation year, after constructing the supported, mixed target policy π_1 and the masked baseline π_0^{supp}, the doubly robust (DR) and self-normalized doubly robust (SNDR) estimator values are computed for each and form the paired improvement:

$$\Delta = V(\pi_1) - V\left(\pi_0^{\text{supp}}\right) \qquad (3-8)$$

To quantify uncertainty in policy values, a paired, spatial cluster bootstrap is used to obtain a point estimate and a 95% CI for Δ. Using 1000 replicates, an empirical distribution of Δ is formed. The 2.5th and 97.5th percentiles of this distribution give a 95% confidence interval for the policy's expected net profit improvement.

Clusters are defined by rounding latitude/longitude to fixed-precision tiles (≈1 km); all records in a tile move together. For each bootstrap replicate, tiles are sampled with replacement, rebuild the index set, and recompute the statistic. This spatial clustering preserves within-tile correlation and avoids the underestimation of uncertainty that would occur with IID-by-field resampling. The resulting CIs reflect both model variability and finite-support uncertainty under the logged behavior.

A handful of policies are evaluated under the profit objective, sweeping hyperparameters λ and ε, and the one with the highest LCB for Δ_P is chosen, requiring at a minimum that the LCB of the paired CI must exceed zero to demonstrate a statistically significant uplift. Yield is treated as a descriptive secondary outcome: under the same π_1, trimmed subset, and weights, the auxiliary yield surrogate model $\hat{v}$ expectations $V_Y(\pi_1)$ and $V_Y(\pi_0^{\text{supp}})$ are computed to isolate agronomic response from price effects.

In conclusion, the policy evaluation methodology rigorously tests the learned fertilizer policy against the historical data using advanced OPE metrics and diagnostics. This approach guards against model optimism and verifies that the policy indeed offers a potential benefit. Using DR and SNDR estimation and bootstrap confidence intervals

grounds the evaluation in established statistical techniques, lending credibility to the findings. All evaluations are presented in Chapter 4 (Results).

3.8 Decision Support Tool Development and Interface

The embedding of the offline evaluated policy into a decision support tool, to close the loop from data science to actionable fertilizer recommendations, is a key component of this praxis. This section describes how the learned policy is converted into a user-facing web application and explains the interface design and safeguards included to make the recommendations both useful and trusted by farmers or extension agents.

Tool Overview. The tool that implements the contextual bandit policy for fertilizer recommendations is developed as a prototype bilingual web application. The app is intended for use by agronomists in the field or farmers who have access to a computer. Users provide a set of pre-plant inputs, and the tool returns an N–P_2O_5–K_2O recommendation along with predicted profit and yield outcomes and a side-by-side comparison to the behavior policy's most likely supported action for the given context.

The app initially collects farm-specific conditions from the user: *latitude*, *longitude*, *elevation*, *slope*, *clay percentage*, *CEC*, *soil organic matter*, and *soil pH*, plus the *season year* and *planting date*. Management conditions, represented by the *system* and *tillage* categories, are selected and internally one-hot encoded for the farm's context feature vector. The user also inputs weather values across six pre-plant time windows (V1–V6), covering *precipitation*, *solar radiation*, *maximum* and *minimum temperature*, and *vapor pressure*, which are then aggregated internally as described in section §3.5. All

of these entries are combined into the context feature vector X that drives the system's recommendations.

By default, the app assumes the same price vector described in section §3.5, so that profit comparisons are consistent with results of the offline evaluation. Policy construction follows the same conservative design as described in section §3.6. The behavior policy $\pi_0(a \mid x)$ is fitted and masked to historically supported action bins – probabilities of bins with less than 30 counts are set to zero and then the distribution is renormalized. The baseline fertilizer rate is defined in the app as the modal (most common) supported action from the logged behavior policy π_0 for the given context. Next the distribution π_1 is formed by identifying the action maximizing profits within the supported action grid according to the surrogate $\hat{\mu}$, allocating a small exploration mass $\varepsilon = 0.10$ across supported cells and applying no masked-π_0 mixing ($\lambda = 0.0$), to mirror the optimal policy that was validated by OPE. π_1 is then projected under the SPIBB constraint: probability mass on bins with fewer than 60 logged observations is set to zero, except for the baseline bin. An action is then sampled from this projected π_1 distribution. If the sampled bin remains below the 60-count threshold, the app conservatively recommends the baseline action. The UI displays the important policy settings – ε, λ, and the support thresholds – so users can see how the recommendation was formed and constrained.

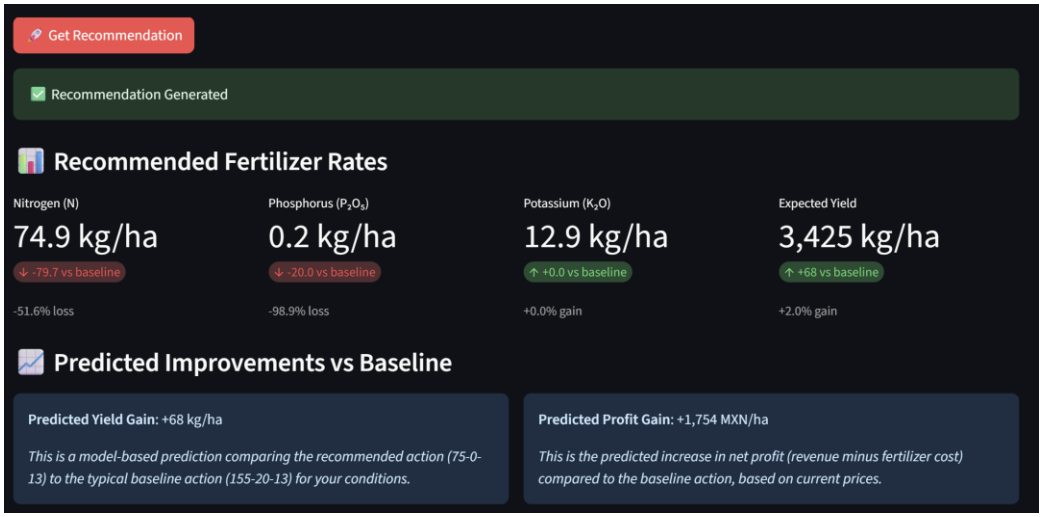

Figure 3-2. Screenshot of the fertilizer recommender app (English)

Each recommendation is displayed together with the formally defined baseline. The main panel then shows the recommended N, P_2O_5, and K_2O rates in kg/ha, the predicted yield, predicted profit, and the differences relative to that baseline; percent-change captions are rendered beneath the metrics to clarify the size and direction of changes. All comparisons are model-based predictions and are presented with a clear note that they are not guarantees.

To keep recommendations agronomically reasonable, the app enforces visible guardrails on the suggested rates; the sidebar documents the maximums and an internal check flags whether the chosen action is within those limits before rendering the results to the user. Current caps are $N \leq 240$, $P_2O_5 \leq 90$, and $K_2O \leq 60$ kg/ha. The interface also exposes a simple confidence label for each generated recommendation; the label reflects whether the sampled action lies in historically supported cells under the configured minimum support, or whether the recommendation deferred to the baseline action under the SPIBB support constraint.

Finally, while the current application is browser-based, the same recommendation workflow can be delivered through lightweight channels to improve accessibility for farmers who may not regularly use a computer. The inputs and outputs described above map naturally to SMS or WhatsApp messages, allowing a user to send minimal and receive a concise N–P_2O_5–K_2O recommendation with predicted yield and profit and a short explanation in reply.

To foster understanding and trust, in the future the app could provide brief reasons behind the recommendation (i.e., "Your soil organic matter is high and you had good pre-season rain, so a moderate N rate is sufficient."), and while the app is featured in Spanish and English, in the future it could also provide support in local native languages. Additionally, and very importantly, the development of such a tool should be informed by stakeholder feedback – by incorporating such feedback, the tool can be better aligned with how decisions are actually made in the field, making it more likely to be adopted.

In summary, the developed decision interface ensures that the contextual bandit policy's output is delivered in a practical and understandable manner. By providing expected outcomes and comparing them to baseline, the comparative advantage of the recommendations are made easy to understand. This approach follows best practices in decision support system design: simplifying complex model predictions into relevant, actionable guidance to facilitate their integration into the decision-making process of end-users.

3.9 Modeling and Experimentation

All experiments in this praxis were implemented in Python using a set of reusable scripts for preprocessing, modelling, off-policy evaluation and deployment. The preprocessing script *preprocess.py* constructs the dataset by loading the Chiapas field trial data, aggregating V1–V6 weather into compact features, excluding outcome proxies and fertilizer actions from the covariate matrix, and retaining only variables knowable at or before planting, including *System* and *Tillage* as categorical context features.

The script writes a compressed NumPy archive containing the context features, three-dimensional fertilizer action array, yield reward, year labels, geo-tile clusters, field identifiers, coordinates and the ordered feature list. Together, this archive provides a consistent input representation for all downstream experiments.

Reward and propensity modelling share a common core module, *core.py*, which utilizes NumPy and pandas together with the gradient-boosting libraries XGBoost, LightGBM and CatBoost, and scikit-learn components such as RobustScaler, mutual_info_regression, RidgeCV and calibration utilities.

Within this module, the AdaptiveBinner class performs discretization of continuous N, P and K rates, forward-year folds and recency weights implement the temporal cross-validation scheme, bootstrap_ci provides non-parametric confidence intervals, and the JointPropensityModel class offers a multiclass XGBoost-based model for the joint N-P–K action distribution, including reliability-curve diagnostics.

As summarized in §3.5, the StackedRewardModel class stacks XGBoost, LightGBM and CatBoost base regressors under a RidgeCV meta-learner on robustly

scaled, mutual-information-filtered features to capture nonlinear fertilizer–response patterns in tabular agronomic data.

XGBoost, LightGBM and CatBoost are widely used open-source gradient-boosting frameworks for structured data, and RobustScaler and mutual_info_regression are standard scikit-learn tools for robust feature scaling and mutual-information-based feature selection. The off-policy evaluation script *ope.py* builds on these components to perform by-year DR and SNDR estimation with overlap trimming, spatial cluster bootstrapping and PSIS weight smoothing, relying on ArviZ's implementation of Pareto-smoothed importance sampling to compute smoothed weights and $\hat{k}$ diagnostics that are then summarized in JSON and CSV result files.

A companion script, *diagnostics.py*, reuses the same AdaptiveBinner, StackedRewardModel and JointPropensityModel classes to compute per-year surrogate and propensity diagnostics and uses Matplotlib to generate reliability and calibration plots and by-year summary tables.

Validated policies and surrogate models are then embedded in the Fertilizer Advisor web application implemented in *app.py* using Streamlit together with NumPy, pandas and joblib for data handling and model loading. Streamlit is an open-source Python framework for turning scripts into interactive web apps, making it well-suited for rapidly prototyping data-driven decision-support tools.

Throughout the project, these Python scripts were developed and iteratively refined in the Visual Studio Code integrated development environment (IDE), a widely used cross-platform source-code editor for Python and data-science workflows, which

provided a practical environment for experimenting with model variants, diagnostics and user-facing interface changes.

3.10 Conclusion

By integrating concepts from machine learning, causal inference, and agronomy, this methodology provides a template for evidence-based decision recommendations in agriculture. The next chapter presents the results of applying these methods, including evaluating the surrogate model's accuracy and the recommended policy's estimated performance gains, thereby demonstrating the effectiveness and limitations of the approach in practice. Throughout, the methodological choices are justified in the context of existing research in offline contextual bandits and agricultural decision support. The rigor of Chapter 3 lays the foundation for the credibility of the results and conclusions in subsequent chapters.

Chapter 4—Results

4.1 Introduction

This chapter reports the predictive and counterfactual results produced by the reward-model ensemble and the off-policy evaluation (OPE) pipeline. It begins with the out-of-fold performance of the stacked reward model, discuss the joint propensity model coverage and calibration, and then turn to the OPE estimates of profit uplift under a conservative acceptance rule that hinges on the lower confidence bound (LCB) of the self-normalized doubly robust (SNDR) estimator. The chapter then summarize by-year policy performance, fertilizer usage, and diagnostics and characteristics of the trimmed subset. Unless otherwise noted, π_0 denotes the support-masked baseline π_0^{supp} (baseline renormalized on supported N–P–K cells). All reported values $V(\cdot)$, improvements Δ, and diagnostics in this chapter are computed on the overlap subset defined by the trimming rules in section §3.5.

4.2 Predictive performance of the reward-model ensemble

The reward model is a stacked ensemble trained in a forward-by-year scheme with site-grouped cross-fitting inside each evaluation year. Year-level performance metrics (R^2 and RMSE) are computed by aggregating fold predictions. More precise methodological details for the surrogate model appear in the previous chapter.

Table 4-1. Performance metrics of surrogate model

Evaluation Year	R^2	RMSE
2013	0.60	3,401
2014	0.75	3,094
2015	0.20	4,939
2016	0.77	2,983
2017	0.70	3,383
2018	0.62	3,730

As seen in Table 4-1, across historical folds, $\hat{\mu}$ is strong in all years with the exception of 2015, a clear outlier, where R^2 falls to 0.20 and RMSE climbs to 4,939 MXN/ha. Excluding this year $\hat{\mu}$ achieves an average R^2 of 0.69 and an average RMSE of 3,318 MXN/ha.

Why is 2015 an outlier? Notably the logged action distribution in 2015 is very different: farmers used much higher fertilizer rates on average (see Table 4-4 below). Farmers' 2015 decisions pushed into higher N–P–K regimes than the model had seen in in prior years, so the reward model had to extrapolate. Because validation on 2015 only uses 2013 and 2014, with recency weighting that down-weights less recent years, the effective training set is narrow; That makes the model sensitive to any 2015 shift in weather, soil conditions, or practice patterns.

Notably, 2015 coincides with a major El Niño event that drove unusual drought and irregular rains across southern Mexico – the kind of mid-season weather the pre-plant feature set can't anticipate. Martín Pérez (2015) documents stunted corn (*maíz enano*) in municipalities like Comitán, La Trinitaria, Las Margaritas, and reports that 53 municipalities in Chiapas were in moderate drought and 5 in severe drought – with approximately 49% of the state being affected. Thus, if fields suffered dry spells after planting this year, the realized profits would be far below what a model trained on earlier years would predict.

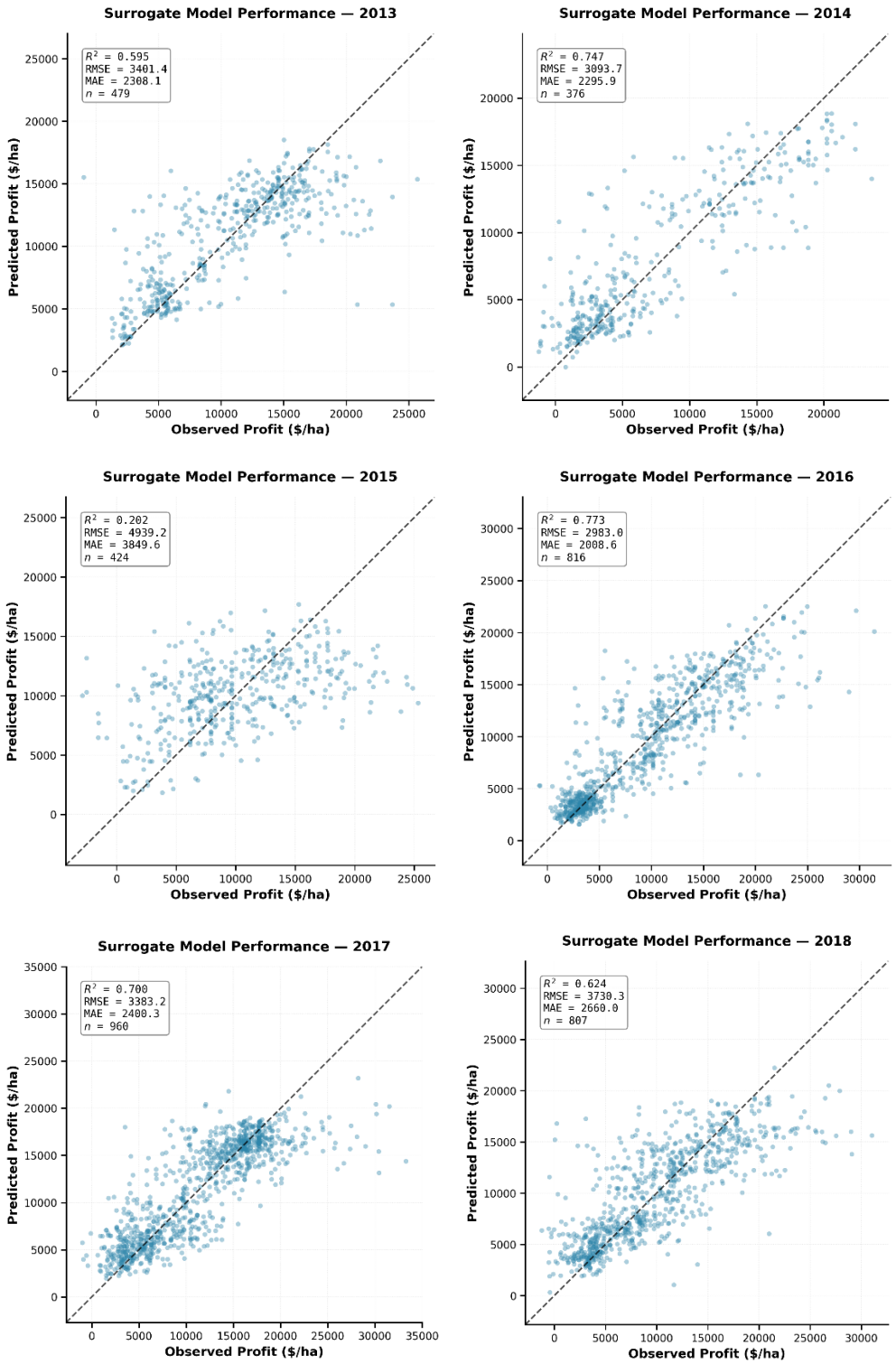

Figure 4-1. Surrogate predicted profits vs. observed profits (2013-2018)

Looking at Figure 4-1, a classic regression-to-the-mean pattern can be observed across evaluation years: overprediction at very low observed profit (points sit above the line) and underprediction at very high profit (points below the line), indicative of a slightly under-confident surrogate – it shrinks extremes toward the center. Nonetheless, the overall level of performance is respectable for agricultural outcomes, which are notoriously noisy and variable year-to-year, but it also leaves room for further refinement in potential future studies.

4.3 Joint Propensity Model Calibration and Diagnostics

For each evaluation year $y \in \{2013, \ldots, 2018\}$ a joint logging-policy model $\pi_0(a \mid x)$ is refit using data from years $< y$, as well as folds other than the test fold for the given year as described in Chapter 3. Actions are then discretized triplets (N, P, K) on the 36-cell grid; an XGBoost multiclass classifier is trained on the compact label set of observed joint cells, probabilities are then expanded back to the full grid; unseen actions remain at exactly zero probability.

Action Coverage. Coverage improves over time. In early years the logging data touch 25–26 of 36 cells; by 2018, 35/36 cells, with 24 cells having at least 30 samples – evidence of broad support in later seasons, which reduces extrapolation risk in OPE. Year-by-year counts are in Table 4-2 below.

Table 4-2. Propensity model action coverage

Year	Cells observed	Cells observed (counts > 30)
2013	26	5
2014	25	9
2015	29	11
2016	32	12
2017	34	20
2018	35	24

Calibration. The propensity estimator was rigorously checked for calibration. Expected Calibration Error (ECE) sits in a tight band of 0.006–0.013 across years (see Figure 4-2); reliability curves for each test year show mild over-confidence overall, especially in the highest bins, but otherwise close alignment between predicted and empirical frequencies. This is deemed to be an acceptable imperfection that does not alter the downstream OPE results, because: the primary estimator is doubly robust and self-normalized (SNDR), and as discussed in Chapter 2, small propensity error is absorbed by the outcome model term ($\frac{1}{n}\sum_{i=1}^{n}$ [$\hat{\mu}\left(x_i, \pi_1(x_i)\right)$ in Equation 2-3). Consistency is retained if either the reward model or the propensity model is well specified, self-normalization dampens variance from the remaining weight noise, and therefore in

practice the SNDR correction substantially damps slight miscalibration effects. Therefore, together with the strict overlap criteria, the subsequent OPE claims are already conservative with respect to estimator noise and small model errors.

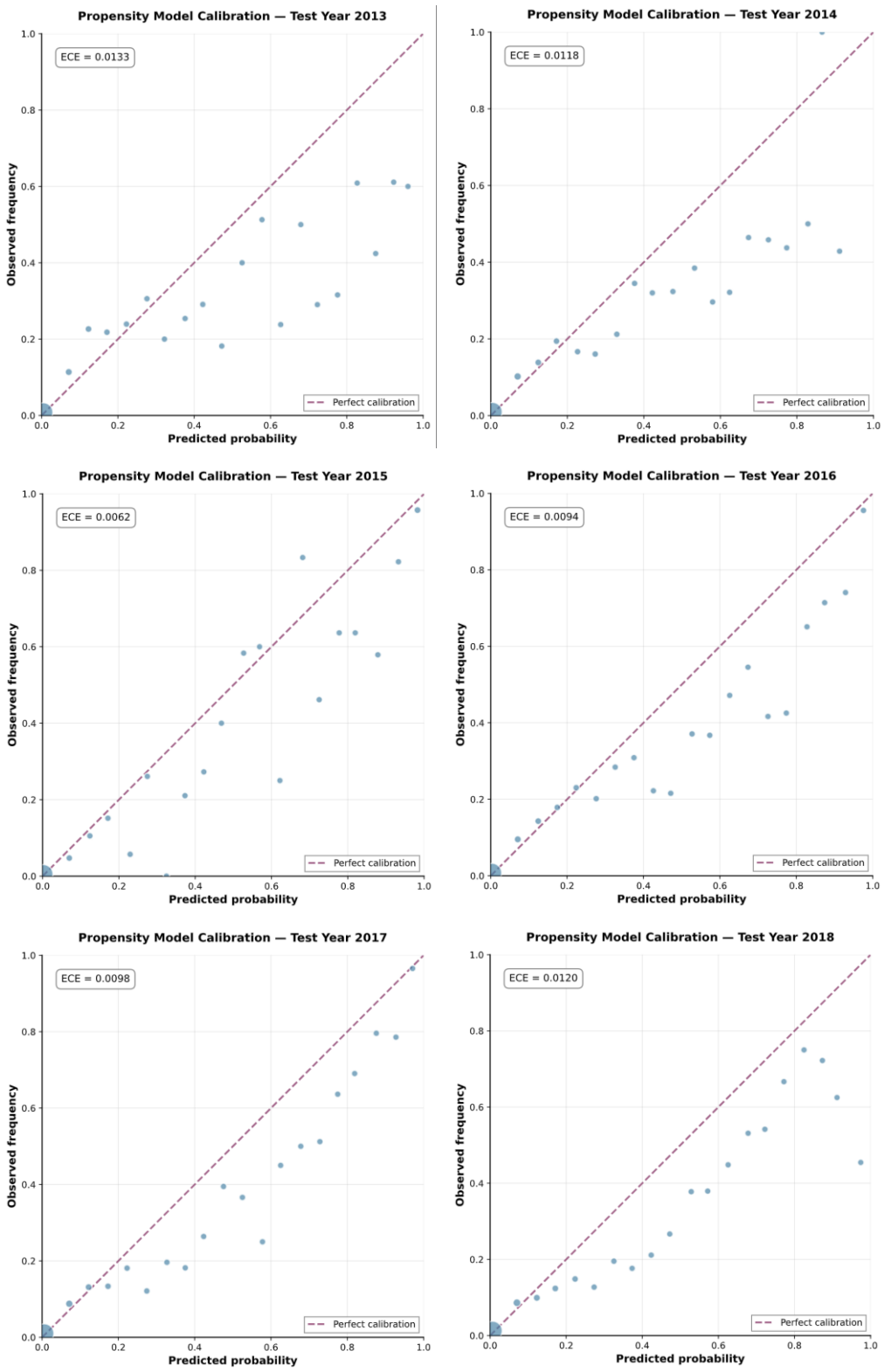

Figure 4-2. Propensity model reliability curves by year (2013-2018)

Propensity tails and overlap. Figure 4-3 shows year-by-year histograms of π_0 on the logged action. The distributions are substantially left-skewed in 2013–2014 and 2017–2018, indicating large mass at small propensities (rare logged actions). This mass near zero motivates the trimming to an overlap subset, that is described in section §3.5.

The outlier year 2015 exhibits a slight left tail but substantial mass at high propensities, meaning farmers' behavior was concentrated in a narrow action band – consistent with a year where patterns are atypical.

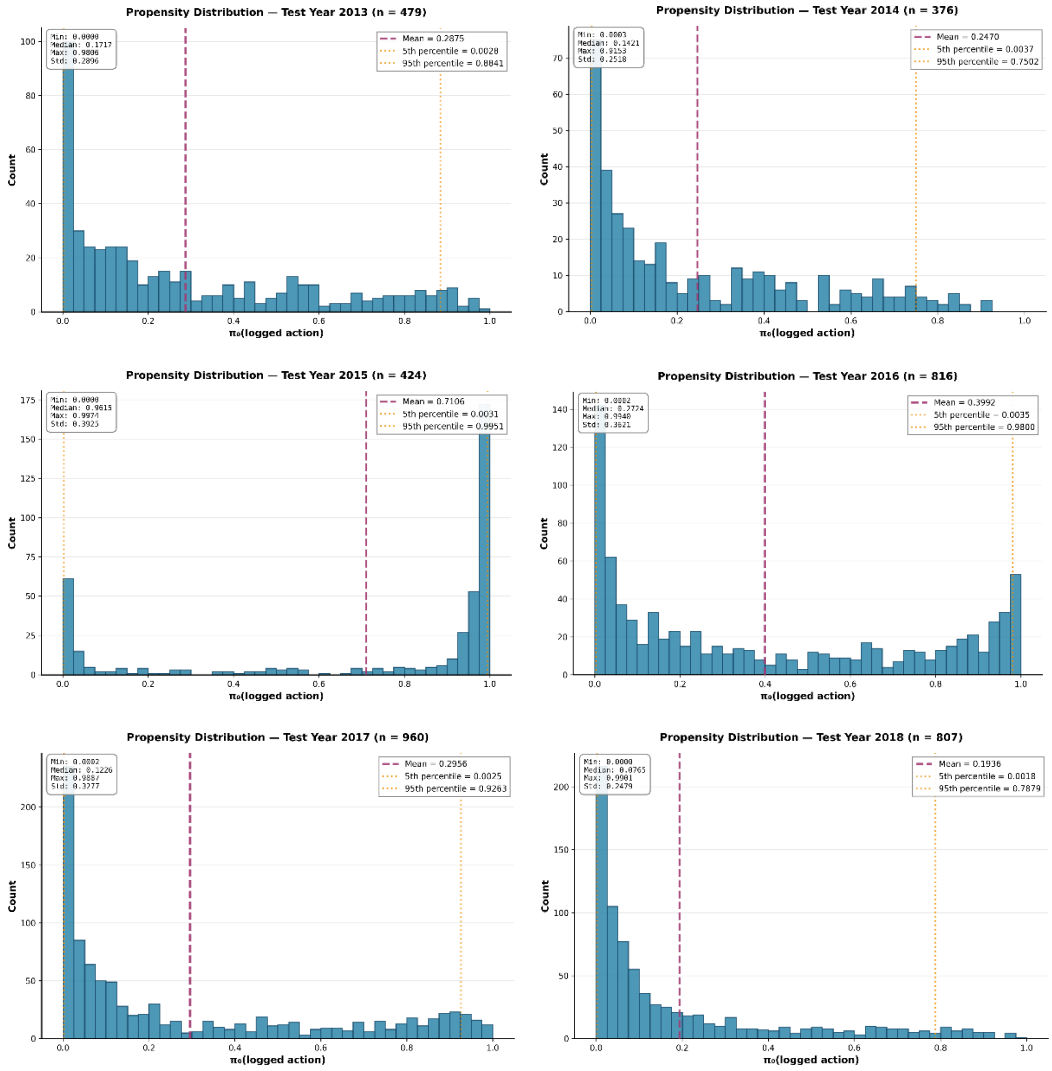

Figure 4-3. Behavior propensities of logged action $\hat{\pi}_0(a_i \mid x_i)$ by year (2013-2018)

In summary, the joint propensity model is accurate and well calibrated, with broad and improving support and controlled variance in importance weights. These properties are essential prerequisites for trusting the off-policy policy evaluation that uses this model's predictions.

4.4 Policy Performance in Offline Evaluation

This section presents the offline evaluation of learned fertilizer recommendation policies using doubly robust (DR) and self-normalized doubly robust (SNDR) estimators. Multiple candidate policies are presented, each with varying degrees of baseline-mixing and exploration, and identify which policy achieves the highest profit gains for farmers, and which estimator provides the most reliable estimates.

These policies vary in their degree of conservatism – some remain closer to farmers' historical practices, by mixing a portion of the logged policy or allowing less exploration, while others are more aggressive in recommending different N–P–K rates. The policies are evaluated on the held-out field data by predicting each field's profit if the policy's recommendation had been followed. Table 4-3 summarizes the best-performing policy's outcome, as estimated by both DR and SNDR. This policy is the one that achieved the highest profit uplift LCB among the candidates without violating the $\hat{k} < 0.7$ diagnostic threshold. The best policy was an aggressive strategy with no baseline-mixing and minimal exploration ($\varepsilon = 0.1$). Crucially, overlap trimming is applied to exclude field observations lacking adequate support (as described below in §4.3), ensuring the evaluation focuses on the overlap subset where the estimates are reliable.

Table 4-3. Offline evaluation of best policy (V_P in MXN/ha, V_Y in kg/ha)

Estimator	$V_P(\pi_0)$	$V_P(\pi_1)$	Profit Gain Δ (95% CI)	$V_Y(\pi_0)$	$V_Y(\pi_1)$	Yield Gain Δ (95% CI)	$\hat{k}$
SNDR	10,238	10,820	+582 (+407 to +776)	3,613	3,811	+199 (+144 to +259)	0.63
DR	10,238	10,857	+619 (+455 to +771)	3,613	3,822	+209 (+157 to +256)	0.63

Both DR and SNDR estimators agree on a significant profit improvement under the policy, but SNDR yields a slightly more conservative point estimate and LCB. For the SNDR estimator, self-normalization introduces a small bias but reduces variance, which on average across policy evaluations results in tighter confidence intervals; however, in the case of the best policy, DR achieved a slightly higher mean gain than SNDR, a tighter confidence interval, as well as a higher LCB (455 vs. 407 MXN/ha). This finding is presented as the main result, and it is noted that the SNDR results are very similar and corroborate the policy's positive impact.

The best policy achieves an overall estimated profit uplift of 582 MXN/ha, a 5.7% increase over the baseline profit of 10,238 MXN/ha, and a corresponding yield uplift of 5.5%. This magnitude of improvement is meaningful in the smallholder context. The results demonstrate that a data-driven fertilizer policy, learned from historical trials and

evaluated with rigorous offline methods, can outperform farmers' current practices with statistical significance.

Although ultimately the no-mixing, minimal-exploration policy (λ=0, ε=0.1) is selected as the primary specification, it is useful to situate it among the other candidates. The sweep covered both conservative and more aggressive recommendation strategies. A clear pattern is observed (to see the offline evaluation results of all policies considered, see Table A-1 in Appendix A): reducing conservatism – i.e., lower baseline-mixing (λ) and lower on-support smoothing (ε) – raised mean profit, consistent with intuition. Notably, the no-mixing, no-exploration strategy (λ=0, ε=0.1) achieved a $\hat{k}$ of 0.78, and hence was rejected as not reliable.

4.5 Performance by Year and Fertilizer Usage

To test robustness across seasons, the best policy is examined for year-by-year performance over evaluation years 2013–2018. Table 4-4 shows sizeable inter-annual heterogeneity: the policy underperforms in 2013 (−232 MXN/ha; −1.7%) and is essentially flat in 2015 (−56 MXN/ha; −0.6%), but delivers clear gains in 2014, 2016, 2017, and 2018—ranging from +467 to +985 MXN/ha (+4.4% to +10.2%). The largest relative improvements occur in years with lower baseline profits (notably 2016 and 2018), consistent with the idea that there is greater scope for improvement when initial performance is weaker.

Table 4-4. Best policy profit performance by year (SNDR, MXN/ha)

Year	$V_P(\pi_0)$	$V_P(\pi_1)$	Profit Gain	Percentage Uplift
2013	13,805	13,574	−232	−1.7%
2014	9,687	10,416	+729	+7.5%
2015	10,150	10,094	−56	−0.6%
2016	9,285	10,092	+807	+8.7%
2017	10,705	11,172	+467	+4.4%
2018	9,639	10,623	+985	+10.2%

Table 4-5 shows that yield gains were observed in every year: +9 kg/ha in 2013 (+0.2%), +459 in 2014 (+13.1%), +167 in 2015 (+4.3%), +186 in 2016 (+5.7%), +143 in 2017 (+3.9%), and +217 in 2018 (+6.5%). Profit responses reflect both revenue from maize yield and input costs. Fertilizer-use patterns help explain these outcomes: historical P and K are low in most years (i.e., mean historical K ≤6 kg/ha in 2014, 2016–2018), and the policy systematically raises P and K while moderately raising N in earlier years and cutting N in later years.

Table 4-5. Best policy yield performance by year (SNDR, kg/ha)

Year	$V_Y(\pi_0)$	$V_Y(\pi_1)$	Yield Gain	Percentage Uplift
2013	4,975	4,984	+9	+0.2%
2014	3,514	3,973	+459	+13.1%
2015	3,841	4,008	+167	+4.3%
2016	3,275	3,461	+186	+5.7%
2017	3,638	3,781	+143	+3.9%
2018	3,319	3,536	+217	+6.5%

Concretely, in 2013 and 2014 the policy raises N, P, and K relative to farmer practice (2013: +49/+17/+11 kg/ha; 2014: +69/+41/+35 kg/ha), but only 2014 shows a strong yield response (+459 kg/ha) that more than covers the added cost; 2013's small yield gain (+9 kg/ha) does not, so profit falls (-1.7%). In 2015, farmers historically applied high rates (148 N, 47 P, 33 K kg/ha), and the policy would have increased yield +167 kg/ha but to little economic benefit (-0.6%). From 2016 onward, the policy shifts posture – combining K and P increases with a leaner N regime (2016: −14 N; 2017: +2 N; 2018: −15 N) to produce modest yield gains and the strongest profit improvements (2016: +8.7%; 2017: +4.4%; 2018: +10.2%). Averaged over all years, the policy implies about +21.5 N, +17.0 P, and +15.5 K kg/ha relative to farmer practice.

Table 4-6. Fertilizer changes by year (π_1 vs historical, kg/ha)

Year	Mean Historical N	Mean Historical P	Mean Historical K	ΔN	ΔP	ΔK
2013	149	40	14	+49	+17	+11
2014	113	21	4	+69	+41	+35
2015	148	47	33	+38	+16	+0
2016	104	16	6	−14	+10	+18
2017	97	19	6	+2	+13	+13
2018	99	19	5	−15	+5	+16

In short, the policy's later-year wins come less from "more fertilizer overall" and more from rebalancing nutrients – shifting away from excess N and toward adequate K and P in a context where K and P have been historically under-applied. This year-specific pattern – consistent yield gains and profit improvements in most seasons – supports the conclusion that the policy's benefits are robust to the temporal variation present in the evaluation data.

4.6 Overlap Diagnostics and Trimming Results

A critical aspect of the evaluation is ensuring that the policy is only credited with improvements where there is sufficient overlap – the data support to make reliable counterfactual predictions. As discussed in Chapter 3 (§3.6), overlap trimming rules are applied to exclude evaluation examples that violated support assumptions. Specifically, any field-case where the logging policy's probability of the historical action ($\pi_0(a|x)$) was below 0.01, or the importance weight ratio w_i exceeded 10, or the action fell in a joint action bin with fewer than 30 samples in the logging policy, was removed from the overlap subset.

Applying these criteria yields an overlap subset that includes 71.3% of 3,862 evaluation rows (2,752 kept; 1,110 trimmed). At the estimator level, weight-tail behavior is well-controlled on the overlap subset: the overall $\hat{k}$ when restricted to this subset is 0.63, compared with 1.36 on the untrimmed full subset – indicating that trimming substantially reduces heavy-tail risk in the importance weights.

Retention varies across seasons but remains broadly representative of the panel. Early-panel scarcity and policy deviations are most binding in 2013, while mid-to-late years exhibit stronger support for the learned recommendations. Geographically, inclusion rates vary meaningfully across space (Figure 4-8); exclusions are not confined to a single location. Nutrient-wise, retention is heterogeneous and driven by support and propensities, but cases with low-N, high-P and K and those with K and no P among the most trimmed combinations (Figures 4-9 and 4-10).

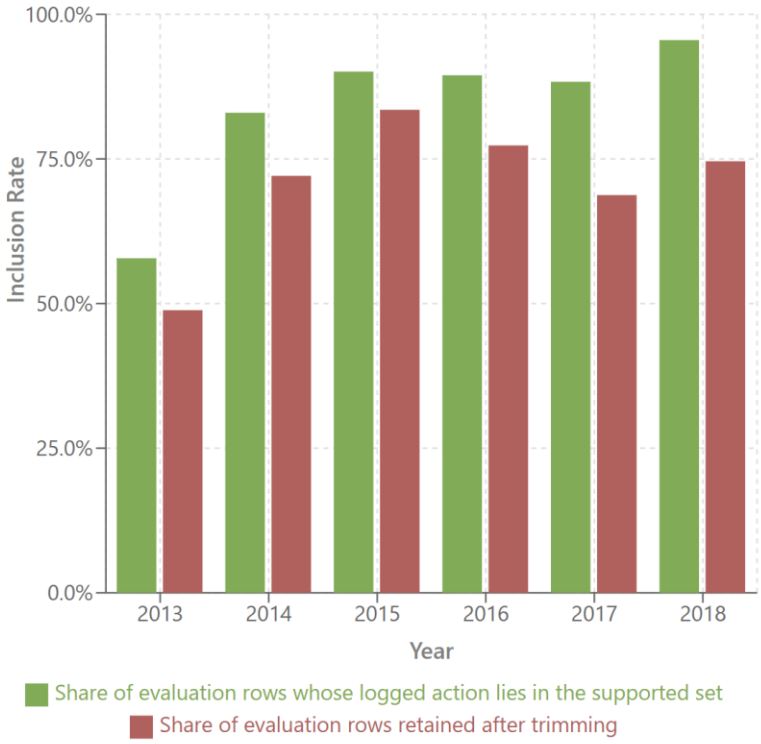

Figure 4-4. Share included in the overlap subset by year

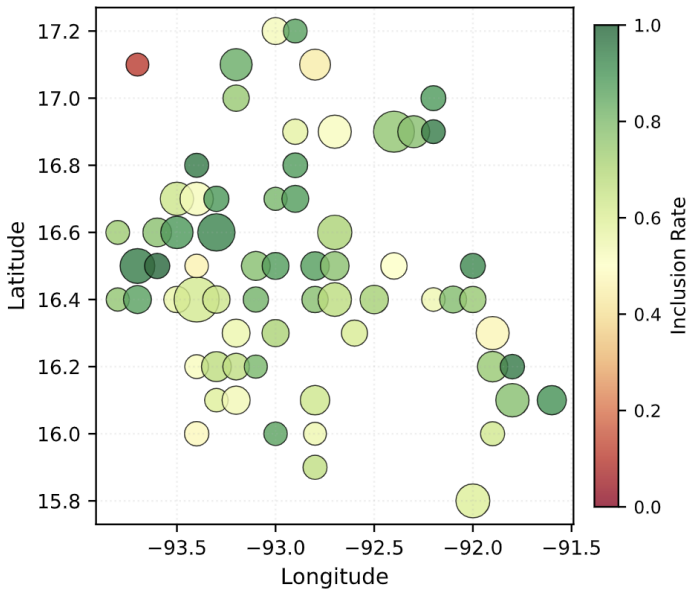

Figure 4-5. Geographic distribution of overlap subset

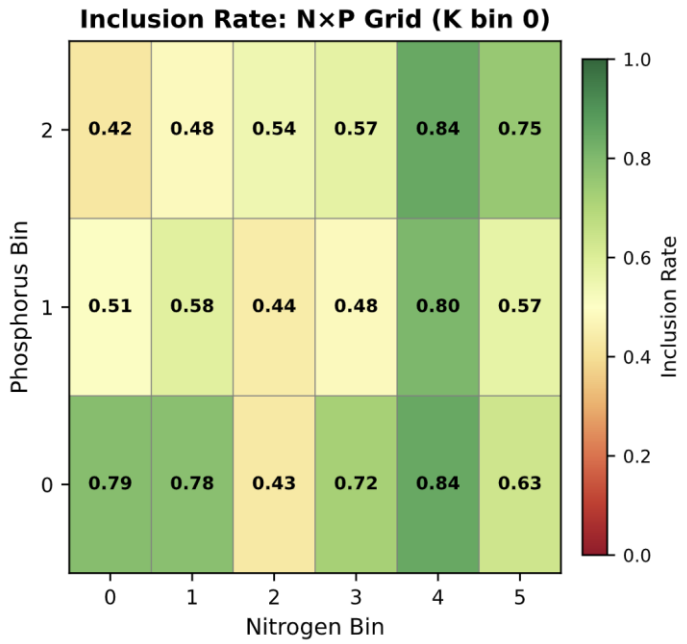

Figure 4-6a. Fertilizer distribution of overlap subset (NxP, K bin 0)

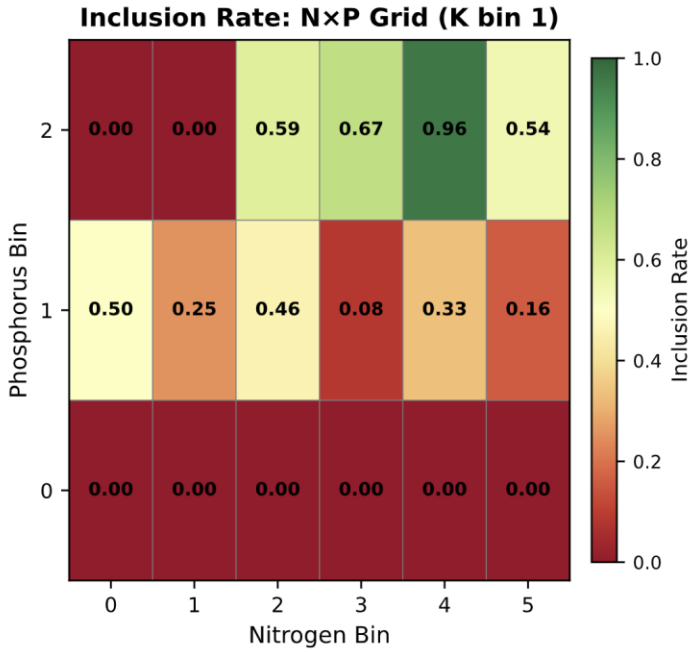

Figure 4-6b. Fertilizer distribution of overlap subset (NxP, K bin 1)

The π_0 floor is the most common signal of non-overlap (around one-fifth of rows are flagged by this condition), reflecting historical cases in which the actions that farmers chose were very unlikely for their given context according to π_0. Low logged-bin support is the next most frequent, accounting for about one-seventh of rows, capturing rare N–P–K action bins globally on the observed side. Extreme importance ratios are comparatively rare, around a few percent, but, when present, are classical markers of extrapolation. These indicators are not mutually exclusive; a given exclusion may trigger multiple reasons.

The net effect of trimming is to anchor evaluation claims to interpolative regions of the Chiapas dataset (Trevisan et al., 2022), where the logged policy meaningfully overlaps the learned target policy in both propensity and support. This is especially important in early seasons with sparser experimentation. The resulting overlap subset (71%) spans all years and management contexts used in the analysis, and it delivers stable, defensible estimates – precisely because the heavy-tail risk observed on the full dataset is curtailed after trimming; overall $\hat{k}$ =0.63 on overlap vs. 1.35 on full. All reported OPE results above therefore pertain to, and should be interpreted within, this overlap regime.

4.7 Conclusion

The rigorous evaluation procedure – doubly robust estimation with self-normalization, coupled with overlap-based trimming and ensuring acceptable tail diagnostics – gives high confidence in the reported performance of the fertilizer recommendation policy. The chosen policy delivers a significant profit improvement for

farmers on average, and this conclusion is supported by multiple estimators and comprehensive diagnostics. This praxis has shown which policy parameters yielded the best results, justified the choice of the DR estimator due to its higher LCB, and demonstrated the necessity of trimming unsupported cases from the dataset. The primary result, a policy that can boost profits by 6% within supported conditions, is a significant and trustworthy finding. Chapter 5

will discuss the significance and practical implications of these results, as well as directions for further research.

Chapter 5—Discussion and Conclusions

5.1 Discussion

This research demonstrates that a contextual bandit approach, validated carefully with offline policy evaluation, can yield fertilizer recommendations that result in credible profit gains for smallholder farmers. The surrogate model effectively predicts farmers' profit, with respectable accuracy given the inherent noise present in agronomic and price data. The 2015 season was a notable outlier, with $R^2 = 0.20$, a symptom of particularly challenging conditions as described in the previous chapter.

In tandem, the joint propensity model accurately learned farmers' historical fertilizer choices. Only 1 of the 36 discrete N–P–K combinations remained unobserved and most cells contained ≥ 30 samples. While the reliability curves indicated mild overconfidence – the overall calibration remained strong (with test-set ECE ranging between 0.006–0.013) and, importantly, does not compromise the off-policy evaluation for the two reasons discussed in Chapter 4. Additionally, the self-normalized doubly robust (SNDR) estimator dampens modest propensity errors through its outcome-model term, preserving consistency when either the reward or propensity model is well specified.

The aggregate economic signal is both positive and statistically precise. Under DR, the best policy attains an estimated profit of 10,857 MXN/ha versus a baseline of 10,238 MXN/ha, a 619 MXN/ha (95% CI: 455, 771); SNDR corroborates with 582 MXN/ha uplift (95% CI: 407, 776). The weight-tail diagnostic on the evaluated overlap-

trimmed subset is well within accepted bounds ($\widehat{k} = 0.63$), supporting estimator stability. These pooled gains mask meaningful inter-annual heterogeneity, however. By-year estimates show underperformance in 2013 (−1.7%), non-significance in 2015 under SNDR (although significant under DR), and statistically significant improvements in 2014, 2016, 2017, and 2018 (7.5% to 10.2%). Such variation is consistent with changing seasonal conditions and price environments as well as the surrogate model's weaker fit in 2015.

What, then, does the policy change in practice? The recommendations do not simply add fertilizer. Instead, they rebalance nutrients in a way that is aligned with documented under-use of P and especially K in this setting. Averaged across years, the policy implies +18.8 kg N, +14.9 kg P_2O_5, and +15.0 kg K_2O relative to farmer practice. Importantly, from 2016 onward the policy reduces N while increasing K, and modestly P, and these rebalanced prescriptions coincide with the largest profit gains. This pattern points to corrections of historical K and P deficits and avoidance of diminishing returns to N as key channels for improvement.

The credibility of these conclusions follows from a deliberately conservative evaluation design. First, claims are restricted to an overlap subset defined by transparent trimming, retaining 71% of evaluation rows and reducing $\hat{k}$ from 1.36 to 0.63. Second, a strictly positive lower 95% confidence bound (LCB > 0) under clustered bootstrap is required to declare per-year improvements, thereby privileging high-confidence, interpolative regions of the data.

Finally, the translation from analysis to decision support preserves these guardrails. The web application replicates the discrete action grid under the same price vector used in OPE, masks to historically supported cells, exposes the same small on-support exploration and no baseline-mixing settings that govern the policy, and enforces fertilizer caps for safety.

Taken together, this study's findings suggest that a conservative contextual bandit policy can yield statistically reliable and practical fertilizer recommendations that nudge fertilizer usage toward more balanced regimes and improved profits. The strength of the approach derives not from aggressive optimization but from respecting overlap, stabilizing weights – an approach that is appropriate for advisory systems targeting smallholders and that provides a defensible baseline for future prospective validation in the field.

Table 5-1 directly compares the data, methods, and results of this praxis with a handful of other SSNM studies. A more detailed account of the praxis's methodology appears in Chapter 3, and the results in Chapter 4.

Table 5-1. Comparison of SSNM methods and results

Study	Data & Context	Methods	Results
Current Praxis	Historical on-farm maize trials in Mexico (Chiapas), per-plot features (soils, weather, management), 2012–2018 (7 seasons).	Framed as a contextual bandit with one pre-season fertilizer decision. Trained a stacked tree-based surrogate; then learned a conservative policy and evaluated with Self-Normalized Doubly-Robust (SNDR) offline evaluation.	+6% profit (+582 MXN/ha); vs farmer practice, with temporal variability (−1.7% in 2013; +10.2% in 2018); +5% yield uplift; variable recommended N-P-K changes vs farmer practice; early years raised all three, while from 2016 onward the policy reduced N (−13%) and increased K and P, corresponding with the years with the strongest profit gains.
Chernet et al. (2024)	Farmer-managed on-farm wheat trials across 277 sites in four Ethiopian districts (Basona Worena, Lemo, Goba, Siyadebir); field validation conducted in 2021.	Machine-learning–generated site-specific fertilizer recommendations; farmer-replicated randomized complete block design; benchmarked against national (NBFR) and local (LBFR) blanket fertilizer recommendations.	+33% profit uplift vs LBFR (+US$580/ha) and +19% vs NBFR (+US$412/ha); performance superior at 72–75% of sites, with expected local variability; +25% yield uplift vs LBFR, +16% vs NBFR.
Chivenge et al. (2021)	Meta-analysis of 61 SSNM studies across 11 countries (maize, rice, wheat), synthesizing yield, fertilizer N, and profitability.	Quantitative meta-analysis examines non-ML rule-based agronomic decision support tools; validation via on-farm trials.	+12% yield and +15% profitability (+$140/ha) vs. farmer practice, with ~10% less fertilizer N applied.
Pasuquin et al. (2014)	Combined crop simulation and on-farm research at ≥65 sites across 13 maize-producing domains in Indonesia, Vietnam, and the Philippines (2004–2008).	SSNM using omission plots / QUEFTS-derived targets; on-farm trials plus simulation and Monte Carlo risk analysis across production and price scenarios.	+1.0 t/ha (+13%) yield vs. farmer practice, average N rate −10% (with more K where deficient), N-use efficiency +42%, profitability +$167/ha per crop (+15% of net return); biggest gains in favorable rainfed areas.
Maertens et al. (2023)	3-year RCT with 792 smallholder households in the maize belt of northern Nigeria; SSNM advice delivered via the Nutrient Expert tool.	Two treatment arms—T1: SSNM information; T2: SSNM + info on maize price distributions/return variability; intent-to-treat estimates, quantile regressions; GHG via IPCC Tier-1	Adoption of good fertilizer management rose sharply (up to +116%); T2 increased nutrient application, yields (+18%), and net revenue (+14%) after two years; NuUE improved; GHG effects mixed—average GHG/ha rose under T2 in year 2 (+148 kg CO_2e/ha, +18%), while emission intensity declined at the upper end (−12% to −17%).
Jat et al. (2018)	3-year farmer-participatory experiment at Taraori, Karnal (NW Indo-Gangetic Plains, India) in a maize–wheat system.	Subplots compare nutrient management – farmer fertilizer practice, recommended dose, and SSNM via Nutrient Expert; outcomes include system productivity, water-use efficiency (WUE) and net returns.	SSNM improved mean system productivity (+13%), WUE (+13%), and net profits (+15%) vs farmer practice.

Dobermann et al. (2002)	179 on-farm experiments from 1997–2000 across eight irrigated rice domains in six Asian countries (China, India, Indonesia, Philippines, Thailand, Vietnam).	SSNM framework using nutrient omission plots to estimate indigenous N, P, K; pre-season NPK needs via decision support (QUEFTS-based) with in-season N adjustments using chlorophyll meter/leaf-color chart; agronomic and economic assessments at field level.	Average yield +11%, N rate −4%, profit +$46/ha per crop (+12%), and fertilizer-N recovery efficiency rising from ~30% → ~40% compared to farmer practice; yield gains ranged 0.1–0.6 t/ha across domains, with strongest profitability in China, southern India, and the Philippines.
Basso et al. (2025)	On-farm analysis of 17 field-years from 13 commercial corn fields in Michigan & Indiana (2021–2023); VRN varied only at the second sidedress; 10,439 grid-cell observations; prescriptions compared: in-season NDVI vs multi-year yield history; gross margins computed with USDA corn & N prices.	Quasi-experimental approach with pseudo-treatments (alignment of applied N with NDVI or YH (crop model-based), spatial linear regression with field/year fixed effects and controls (N rate, NDVI/YH levels, weather & site variables), plus a spatial discontinuity analysis on adjacent "rook" cells with the same N rate but different info source; price-sensitivity checks using 2009–2023 extremes.	Heterogeneous profitability: NDVI–YH gross-margin differences ranged from −$410 to +$350/ha (−14.6% to +5.3%) relative to YH. By year, 2021 favored NDVI (+2.0% to +5.3% in most fields; a few −0.8% to −2.6%), 2022 showed no clear pattern (−0.7% to +0.9%), and 2023 favored YH (−14.6% to −1.1%). Price-sensitivity checks did not flip the qualitative rankings, indicating the two information sources are complementary.
Khakbazan et al. (2021)	Field-scale evaluation in western Canada (Manitoba, Saskatchewan, Alberta). Ten sites from 2014–2016 used for the economic analysis; a broader agronomic dataset covered 27 sites (2014–2017). Management zones were delineated from historical yield maps and soil tests.	Factorial design of three yield-based management zones (low/average/high) × four N rates (0, 50, 100, 150% of recommended), replicated four times per site; linear models/ANOVA assessed yield and net revenue, with additional analyses of N use and NUE.	When pooled across 10 fields, management-zone N delivered +$28 to +$65/ha higher net revenue vs. average-yield/uniform management, with strong cross-site heterogeneity (−$91 to +$352/ha vs. baseline). Applied N under zones was 8% lower on average; under a yield-maximizing scenario, total N was ~18 kg/ha less than uniform application.

5.2 Conclusions

This research achieved its stated objectives and demonstrated the feasibility of offline, data-driven fertilizer optimization for maize smallholders. In summary:

- **Objective 1:** The stacked tree-based surrogate model was successfully trained and validated. It met the predefined targets with out-of-fold R^2 averaging 0.69 and RMSE 3,318 MXN/ha with the exception of one outlier year (2015). The model

reliably predicts profit from context and fertilizer inputs across diverse Chiapas fields.

- **Objective 2:** The joint propensity model attained broad and steadily increasing coverage and good calibration. Nearly all action combinations were observed in the data (35/36 cells by the final evaluation year), and the model's expected calibration error was low (ranging between 0.006–0.013 on hold-out data). These properties allowed stable importance weighting for evaluation.
- **Objective 3:** Using the surrogate and propensity models, a support-constrained contextual-bandit policy is learned. The policy optimization used reward predictions to select profit-maximizing actions, while enforcing that recommendations remain within historically supported fertilizer regimes and exhibit reasonable importance weights.
- **Objective 4:** Conservative off-policy evaluation with DR and SNDR estimators with bootstrap confidence intervals were performed successfully. The learned policy demonstrated a profit uplift of 5.7 percent, with a corresponding yield uplift of 5.5 percent.
- **Objective 5:** The policy was operationalized by integrating it into a bilingual decision support web application. This app accepts farm-specific conditions as input and returning N–P–K recommendations as well as yield and profit estimates, and provides a comparison with the historical baseline so users can gauge the value of the recommendations.

Therefore, each of the research objectives were met satisfactorily. The system predicts profit response, infers farmer propensities, computes safe fertilizer recommendations, validates them rigorously, and packages them for practical use, providing a complete end-to-end solution from farm data to actionable fertilizer advice.

5.3 Contributions to the Body of Knowledge

This study offers several novel contributions at the intersection of agriculture and machine learning:

- **Advancing data-driven agronomy:** The praxis demonstrates for the first time, so far as can be determined, an effective end-to-end application of offline contextual-bandit learning and causal inference techniques in a realistic smallholder farming context. Unlike prior work that often focuses on yield forecasting, RL simulation using crop models, or on-farm experiments, the approach moves from prediction to offline empirically-vetted policy optimization. By leveraging a rich historical dataset (2012–2018 Chiapas trials) and using advanced ML and causal inference techniques, this praxis demonstrates it is possible to derive trustworthy fertilizer strategies without new field experiments.
- **Safe offline reinforcement learning methodology:** The incorporation of conservative overlap criteria into policy evaluation contributes a replicable framework for robust OPE in agronomy. The use of doubly robust (DR) and self-normalized doubly-robust (SNDR) estimators, cluster bootstrapping, and strict overlap criteria sets a strong standard for reliability. This methodology extends

recent safe RL practices to the agricultural domain and provides a blueprint for other researchers aiming to avoid false positives in offline policy recommendations.

- **Site-specific nutrient optimization insights:** This praxis yields agronomic insights into smallholder fertilizer response. This study finds that modest decreases in nitrogen, and small adjustments in phosphorous and potassium, dependent upon local field conditions, can translate reliably into profit gains, whereas excessive fertilizer typically results in diminishing returns. This pattern validates the understanding of maize response in tropical smallholder systems and confirms the idea that blanket recommendations overlook field-level heterogeneity. This study thus contributes to the literature on site-specific nutrient management (SSNM) by quantifying optimal fertilizer rates under specific conditions.

- **Decision-support tool for impact:** Finally, the development of the bilingual web app represents a contribution to technology transfer. While many studies stop at model development, this study demonstrates how to deliver trustworthy advice to users. The app's design – from input forms mirroring farmers' specific contexts to confidence labels and guardrails – follows best practices in user-centered decision support. This prototype paves the way for scalable dissemination of precision recommendations (via accessible channels such as SMS and WhatsApp), illustrating how machine learning research can directly inform extension services.

This praxis bridges multiple disciplines. It extends predictive ML and offline RL into agronomy, applies causal inference principles, and addresses decision support

deployment concerns. The findings contribute both methodologically and empirically to the growing field of data-driven agriculture.

5.4 Recommendations for Future Research

Building on these results, several avenues are examined that could deepen and broaden the impact of this work.

Field validation trials. The ultimate test is in-situ experimentation. Future work should conduct randomized controlled trials (RCTs) or A/B field studies comparing the policy's recommendations to current practice or local guidelines. Empirical trials would verify that the offline-estimated gains materialize under real-world conditions, and would reveal any unmodeled factors affecting outcomes, such as pest pressure or other undocumented management choices. Gathering post-deployment feedback would also allow retraining the surrogate model on new data.

Price and climate robustness. The surrogate profit model assumes fixed prices (N=16, P_2O_5=12, K_2O=8 MXN/kg; maize 3.5 MXN/kg) across all years. In reality, fertilizer and maize prices fluctuate. The policy's advantage depends on low-enough input costs and high-enough maize prices. Future research should incorporate temporal price fluctuations into the optimization objective. In addition, modeling seasonal forecasts into the policy could improve resilience to weather shocks.

Data and model enhancements. Future research might include improving on both the model and data sets used for modelling. There are a number of ways that additional data can be included to provide better context for the surrogate crop model –

one way could be through the inclusion of remote sensing indexes such as the normalized difference vegetation index (NDVI). This would allow for the enrichment of the context feature set. The addition of more years, more locations, and more management variables to the dataset could also increase the generalizability of the model. Furthermore, using other ML architectures such as deep neural networks could potentially capture complex relationships that may have been missed by trees, although care should be used to maintain interpretability.

Dynamic decision-making. The current formulation treats fertilizer choice as a one-shot decision. In practice as discussed before, farmers may apply fertilizer in stages and respond to mid-season information, adjusting with a top-dress a few weeks after planting. Extending the framework to a multi-stage RL framework could capture these dynamics. Such extensions, however, would require data at each fertilizer application stage, which could be helped by utilizing additional crop monitoring infrastructure.

Human-centered design. The prototype app could be refined through participatory design. Future work should engage smallholder farmers and extension agents in usability testing to ensure that the tool's recommendations and explanations resonate with users. As noted in Chapter 3, adding simple justifications and supporting local languages (Tzeltal and Tzotzil are the most spoken indigenous languages in Chiapas, for example) would increase trust. In addition, trials using accessible communication channels like SMS or WhatsApp could be piloted to evaluate accessibility, message clarity, and adoption barriers. User feedback could also guide adjusting the support thresholds to match farmer risk preferences.

Cross-region generalization. Applying this approach to other crops and regions would validate its generalizability. For example, adapting the framework to wheat or rice, or to smallholder systems in Africa or Asia, could reveal how well the methodology transfers. Comparative studies could highlight differing fertilizer dynamics between regions. Such work would benefit from collaborations with local agricultural research centers and would most likely involve retraining models on region-specific datasets.

Sustainability assessments. Although fertilizer changes are documented under the policy, future work could further evaluate environmental outcomes. Since the policy tends to reduce over-application, it likely lowers excess nitrogen runoff and greenhouse emissions, but quantifying this further would be valuable. Soil health monitoring under the new recommendations could demonstrate broader impacts. Integrating other objectives, like the carbon footprint or biodiversity impact, into the optimization via multi-objective RL is another potential research direction.

Methodological innovations. There are a number of methods on the algorithm side that could reduce the conservativeness of the approach while preserving safety, such as making use of Bayesian policy evaluation or risk sensitive criteria. Additionally, moving away from discretizing the action space to either continuous action spaces or using non-parametric propensity estimation methods could increase the degree of flexibility in the approach. Techniques from robust machine learning (i.e., adversarial re-weighting, domain adaptation), can also be studied to help the policy deal with outliers out of distribution input and rare events.

Agentic AI. As the field of AI and ML moves toward agentic AI – AI that can independently reason, plan, and take action – there is a natural path from the work

presented here to safely-scoped autonomy in agronomic recommendation workflows. The focus of this praxis on causal framing, conservative policy learning, and the use of offline evaluation has already provided important foundational elements for safe agency: OPE using methods like DR and SNDR to estimate counterfactuals, and policy learning that will respect support constraints, which can reduce the risk of unsupported extrapolation in environments where potential users' livelihoods are at stake. Future development of this approach could include richer structural causal models to test the degree of identifiability and distribution shift, couple OPE with cautious online adaptations under pessimistic or uncertainty-based penalty functions, and ensure that the eventual deployment of an autonomous or semi-autonomous recommender system aligns with emerging governance and verification standards and best practices for agentic systems such as traceability, fail-safe mechanisms, and human override gates.

In conclusion, this praxis opens multiple pathways for future work. By integrating additional data, refining models, and engaging end-users, future research can build on the current findings to create even more reliable, effective, and scalable decision-support systems for smallholder farmers. With continued adaptation to diverse crops, contexts, and geographies, the framework aims to ensure that data-generating communities can continue to secure credible improvements in productivity, profitability, and environmental stewardship.

References

Abera, W., Tamene, L., Tesfaye, K., Jiménez, D., Dorado, H., Erkossa, T., Kihara, J., Ahmed, J. S., Amede, T., & Ramirez-Villegas, J. (2022). A data-mining approach for developing site-specific fertilizer response functions across the wheat-growing environments in Ethiopia. *Experimental Agriculture, 58*, e9. https://doi.org/10.1017/S0014479722000047

An, G., Moon, S., Kim, J.-H., & Song, H. O. (2021). Uncertainty-based offline reinforcement learning with diversified Q-ensemble. In M. A. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. Wortman Vaughan (Eds.), *Advances in Neural Information Processing Systems* (Vol. 34, pp. 7436–7447). Curran Associates, Inc.

Azzari, G., Jin, Z., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment, 228*, 115–128. https://doi.org/10.1016/j.rse.2019.04.016

Balemi, T., & Rurinda, J. (2021). Site-specific nutrient management, using Nutrient Expert tool, improved farmers' maize grain yield in Oromia region. *Ethiopian Journal of Crop Science, 8*(1), 25–37. https://www.researchgate.net/publication/348919604

Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially

variable field. *European Journal of Agronomy, 35*, 215–222. https://doi.org/10.1016/j.eja.2011.06.004

Bhat, S. A., Qadri, S. A. A., Dubbey, V., Sofi, I. B., & Huang, N.-F. (2024). Impact of crop management practices on maize yield: Insights from farming in tropical regions and predictive modeling using machine learning. *Journal of Agriculture and Food Research, 18*(3), 101392. https://doi.org/10.1016/j.jafr.2024.101392

Birner, R., Davis, K. E., Pender, J. L., Nkonya, E. M., Anandajayasekeram, P., Ekboir, J. M., Mbabu, A. N., Spielman, D. J., Horna, D., Benin, S., & Cohen, M. (2009). From best practice to best fit: A framework for designing and analyzing pluralistic agricultural advisory services worldwide. *The Journal of Agricultural Education and Extension, 15*(4), 341–355. https://doi.org/10.1080/13892240903309595

Breiman, L. (1996). Stacked regressions. *Machine Learning, 24*(1), 49–64. https://doi.org/10.1007/BF00117832

Cai, S., Zhao, X., & Yan, X. (2025). Towards precise nitrogen fertilizer management for sustainable agriculture. *Earth Critical Zone, 2*, 100026. https://doi.org/10.1016/j.ecz.2025.100026

Campolo, J., & Lobell, D. (2022). Evaluating maize yield response to fertilizer and soil in Mexico using ground and satellite approaches. *Field Crops Research, 276*, 108393. https://doi.org/10.1016/j.fcr.2021.108393

Che, F. (2025). A tutorial: An intuitive explanation of offline reinforcement learning theory (arXiv:2508.07746) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2508.07746

Chen, T., & Guestrin, C. (2016, August). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chernet, M., Liben, F. M., Abera, W., Kihara, J., Wolde-Meskel, E., Tamene, L., Thuita, M., Seleshi, Y., Kebede, A., Erenstein, O., & Rahut, D. B. (2024). Site-specific fertilizer recommendation using data-driven machine learning enhanced wheat productivity and resource use efficiency. *Field Crops Research, 313*, 109413. https://doi.org/10.1016/j.fcr.2024.109413

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal, 21*(1), C1–C68. https://doi.org/10.1111/ectj.12097

Chivenge, P., Saito, K., Bunquin, M. A., Sharma, S., & Dobermann, A. (2021). Co-benefits of nutrient management tailored to smallholder agriculture. *Global Food Security, 30*, 100570. https://doi.org/10.1016/j.gfs.2021.100570

Chua, K., Calandra, R., McAllister, R., & Levine, S. (2018). Deep reinforcement learning in a handful of trials using probabilistic dynamics models (PETS). In *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc.

Corrales, D. C., Schoving, C., Raynal, H., Debaeke, P., Journet, E.-P., & Constantin, J. (2022). A surrogate model based on feature selection techniques and regression learners to improve soybean yield prediction in southern France. *Computers and*

Electronics in Agriculture, 192, 106578. https://doi.org/10.1016/j.compag.2021.106578

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika, 96*(1), 187–199. https://doi.org/10.1093/biomet/asn055

Cunha, R. L. de F., Silva, B., & Avegliano, P. B. (2023). A comprehensive modeling approach for crop yield forecasts using AI-based methods and crop simulation models [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2306.10121

Davis, K., Babu, S. C., & Ragasa, C. (2020). *Agricultural extension: Global status and performance in selected countries*. International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/9780896293755

Deb, R., Ghavamzadeh, M., & Banerjee, A. (2024, December 9). Conservative contextual bandits: Beyond linear representations. arXiv. https://doi.org/10.48550/arXiv.2412.06165

Dougherty, J., Kohavi, R., & Sahami, M. (1995, July). Supervised and unsupervised discretization of continuous features. In P. E. C. I. Prieditis & S. J. Russell (Eds.), *Proceedings of the 12th International Conference on Machine Learning* (pp. 194-202). Morgan Kaufmann.

Droutsas, I., Challinor, A. J., Deva, C. R., & Wang, E. (2022). Integration of machine learning into process-based modelling to improve simulation of complex crop responses. *in silico Plants, 4*(2), diac017. https://doi.org/10.1093/insilicoplants/diac017

Dudík, M., Langford, J., & Li, L. (2011). Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning* (pp. 1097–1104). Omnipress.

Fonteyne, S., Castillo Caamal, J. B., Lopez-Ridaura, S., Van Loon, J., Espidio Balbuena, J., Osorio Alcala, L., Martínez Hernández, F., Odjo, S., & Verhulst, N. (2023). Review of agronomic research on the milpa, the traditional polyculture system of Mesoamerica. *Frontiers in Agronomy, 5*, 1115490. https://doi.org/10.3389/fagro.2023.1115490

Gao, J., Zeng, W., Ren, Z., Ao, C., Lei, G., Gaiser, T., & Srivastava, A. K. (2023). A fertilization decision model for maize, rice, and soybean based on machine learning and swarm intelligent search algorithms. *Agronomy, 13*(5), 1400. https://doi.org/10.3390/agronomy13051400

Gautron, R., Baudry, D., Adam, M., Falconnier, G. N., & Corbeels, M. (2022). Towards an efficient and risk-aware strategy for guiding farmers in identifying best crop management. *arXiv* (arXiv:2210.04537) [Preprint]. https://doi.org/10.48550/arXiv.2210.04537

Gautron, R., Baudry, D., Adam, M., Falconnier, G. N., Hoogenboom, G., King, B., & Corbeels, M. (2024, June 19). Risk-aware bandits for best crop management. In *ARLET 2024: 38th Workshop on Aligning Reinforcement Learning Experimentalists and Theorists (ICML 2024)* [Poster]. OpenReview. https://openreview.net/forum?id=a0HN4QtznG

Gautron, R., Padrón, E. J., Preux, P., Bigot, J., Maillard, O.-A., & Emukpere, D. (2022). gym-DSSAT: A crop model turned into a reinforcement learning environment (arXiv:2207.03270) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2207.03270

Giannarakis, G., Sitokonstantinou, V., Lorilla, R. S., & Kontoes, C. (2022). Personalizing sustainable agriculture with causal machine learning. In *Tackling Climate Change with Machine Learning: Workshop at NeurIPS 2022*. Climate Change AI. (Workshop paper). Retrieved from https://climate-change-ai.s3.us-east-1.amazonaws.com/papers/neurips2022/112/paper.pdf

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Jagerman, R., Markov, I., & de Rijke, M. (2020). Safe exploration for optimizing contextual bandits. *ACM Transactions on Information Systems, 38*(3), 1–23. https://doi.org/10.1145/3385670

Jat, R. D., Jat, H. S., Nanwal, R. K., Yadav, A. K., Bana, A., Choudhary, K. M., Kakraliya, S. K., Sutaliya, J. M., Sapkota, T. B., & Jat, M. L. (2018). Conservation agriculture and precision nutrient management practices in maize–wheat system: Effects on crop and water productivity and economic profitability. *Field Crops Research, 222*, 111–120. https://doi.org/10.1016/j.fcr.2018.03.025

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT

cropping system model. *European Journal of Agronomy, 18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Kakimoto, S., Mieno, T., Tanaka, T. S. T., & Bullock, D. S. (2022). Causal forest approach for site-specific input management via on-farm precision experimentation. *Computers and Electronics in Agriculture, 199*, 107164. https://doi.org/10.1016/j.compag.2022.107164

Kanthilanka, H., Ramilan, T., Farquharson, R. J., & Weerahewa, J. (2023). Optimal nitrogen fertilizer decisions for rice farming in a cascaded tank system in Sri Lanka: An analysis using an integrated crop, hydro-nutrient and economic model. *Agricultural Systems, 207*, 103628. https://doi.org/10.1016/j.agsy.2023.103628

Kazerouni, A., Ghavamzadeh, M., Abbasi-Yadkori, Y., & Van Roy, B. (2017). Conservative contextual linear bandits. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 3913–3922). Curran Associates, Inc.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 3146–3154).

Khakbazan, M., Moulin, A., & Huang, J. (2021). Economic evaluation of variable rate nitrogen management of canola for zones based on historical yield maps and soil test recommendations. *Scientific Reports, 11*(1), 4439. https://doi.org/10.1038/s41598-021-83917-3

Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science, 10*, 621. https://doi.org/10.3389/fpls.2019.00621

Kidambi, R., Rajeswaran, A., Netrapalli, P., & Joachims, T. (2020). MOReL: Model-based offline reinforcement learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33). Curran Associates, Inc.

Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2018). The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine, 24*(11), 1716–1720. https://doi.org/10.1038/s41591-018-0213-5

Kumar, A., Zhou, A., Tucker, G., & Levine, S. (2020). Conservative Q-learning for offline reinforcement learning. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*. Curran Associates, Inc.

Laroche, R., Trichelair, P., & Tachet des Combes, R. (2019). Safe policy improvement with baseline bootstrapping. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 97, pp. 3652–3661). PMLR. https://proceedings.mlr.press/v97/laroche19a.html

Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)* (pp. 661–670). ACM. https://doi.org/10.1145/1772690.1772758

Li, L., Chu, W., Langford, J., & Wang, X. (2011). Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM '11)* (pp. 297–306). Association for Computing Machinery. https://doi.org/10.1145/1935826.1935878

Maertens, M., Oyinbo, O., Abdoulaye, T., & Chamberlin, J. (2023). Sustainable maize intensification through site-specific nutrient management advice: Experimental evidence from Nigeria. *Food Policy, 121*, 102546. https://doi.org/10.1016/j.foodpol.2023.102546

Martín Pérez, F. (2015, August 29). *Sequía. Crece maíz enano ante falta de lluvias*. El Universal. https://www.eluniversal.com.mx/articulo/estados/2015/08/29/sequia-crece-maiz-enano-ante-falta-de-lluvias/

Martínez, F. B., Guevara, F., Aguilar, C. E., Pinto, R., La O, M. A., Rodríguez, L. A., & Aryal, D. R. (2020). Energy and economic efficiency of maize agroecosystem under three management strategies in the Frailesca, Chiapas (Mexico). *Agriculture, 10*(3), 81. https://doi.org/10.3390/agriculture10030081

Moothedath, S., Lee, X. Y., Jubery, T., Ganapathysubramanian, B., & Sarkar, S. (2021, November). A conservative stochastic contextual bandit based framework for farming recommender systems. In *AI for Agriculture and Food Systems Workshop (35th AAAI Conference on Artificial Intelligence)*. (Workshop presentation).

Natchev, V. (2024, November 12). Harnessing AI to empower smallholder farmers: Bridging the digital divide for sustainable growth. *Harvard ALI Social Impact*

Review. https://www.sir.advancedleadership.harvard.edu/articles/harnessing-ai-empower-smallholder-farmers-bridging-digital-divide-sustainable-growth

Orchardson, E. (2019, January 15). Reducing high yield gaps with decision-support apps. CIMMYT. https://www.cimmyt.org/news/reducing-high-yield-gaps-with-decision-support-apps/

Osband, I., Blundell, C., Pritzel, A., & Van Roy, B. (2016). Deep exploration via bootstrapped DQN. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 29, pp. 3003–3011). Curran Associates, Inc.

Pasuquin, J. M., Pampolino, M. F., Witt, C., Dobermann, A., Oberthür, T., Fisher, M. J., & Inubushi, K. (2014). Closing yield gaps in maize production in Southeast Asia through site-specific nutrient management. *Field Crops Research, 156*, 219–230. https://doi.org/10.1016/j.fcr.2013.11.016

Pewsey, A., & García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*, 30(1), 1–58.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (Vol. 31, pp. 6638–6648). https://papers.neurips.cc/paper/7898-catboost-unbiased-boosting-with-categorical-features.pdf

Quan, Z., Zhang, X., Davidson, E. A., Zhu, F., Li, S., Zhao, X., Chen, X., Zhang, L.-M., He, J.-Z., Wei, W., & Fang, Y. (2021). Fates and use efficiency of nitrogen

fertilizer in maize cropping systems and their responses to technologies and management practices: A global analysis of field ^{15}N tracer studies. *Earth's Future, 9*(5), e2020EF001514. https://doi.org/10.1029/2020EF001514

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Rehill, P. (2025). How do applied researchers use the causal forest? A methodological review. *International Statistical Review, 93*(2), 288–316. https://doi.org/10.1111/insr.12610

Sachdeva, N., Su, Y., & Joachims, T. (2020). Off-policy bandits with deficient support. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 965–975). ACM. https://doi.org/10.1145/3394486.3403139

Saito, Y., Udagawa, T., Kiyohara, H., Mogi, K., Narita, Y., & Tateno, K. (2021). Evaluating the robustness of off-policy evaluation. In *Proceedings of the 15th ACM Conference on Recommender Systems (RecSys '21)* (pp. 114–123). Association for Computing Machinery. https://doi.org/10.1145/3460231.3474245

Sapkota, T. B., Jat, M. L., Rana, D. S., Khatri-Chhetri, A., Jat, H. S., Bijarniya, D., Sutaliya, J. M., Kumar, M., Singh, L. K., Jat, R. K., Kalvaniya, K., Prasad, G., Sidhu, H. S., Rai, M., Satyanarayana, T., & Majumdar, K. (2021). Crop nutrient management using Nutrient Expert improves yield, increases farmers' income and reduces greenhouse gas emissions. *Scientific Reports, 11*(1), 1564. https://doi.org/10.1038/s41598-020-79883-x

Shahhosseini, M., Hu, G., & Archontoulis, S. V. (2020). Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science, 11*, 1120. https://doi.org/10.3389/fpls.2020.01120

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports, 11*, 1606. https://doi.org/10.1038/s41598-020-80820-1

Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion, 81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

Siedliska, A., Baranowski, P., Pastuszka-Woźniak, J., Zubik, M., & Krzyszczak, J. (2021). Identification of plant leaf phosphorus content at different growth stages based on hyperspectral reflectance. *BMC Plant Biology, 21*, 28. https://doi.org/10.1186/s12870-020-02807-4

Smidt, H. J., & Jokonya, O. (2022). Factors affecting digital technology adoption by small-scale farmers in agriculture value chains (AVCs) in South Africa. *Information Technology for Development, 28*(3), 558–584. https://doi.org/10.1080/02681102.2021.1975256

Stewart, W. M., Dibb, D. W., Johnston, A. E., & Smyth, T. J. (2005). The contribution of commercial fertilizer nutrients to food production. *Agronomy Journal, 97*(1), 1–6. https://doi.org/10.2134/agronj2005.0001

Su, Y., Dimakopoulou, M., Krishnamurthy, A., & Dudík, M. (2020). Doubly robust off-policy evaluation with shrinkage. In H. Daumé III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning*

(Proceedings of Machine Learning Research, Vol. 119, pp. 9167–9176). PMLR. https://proceedings.mlr.press/v119/su20a.html

Swaminathan, A., & Joachims, T. (2015). The self-normalized estimator for counterfactual learning. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28, pp. 3231–3239). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/hash/39027dfad5138c9ca0c474d71db915c3-Abstract.html

Thomas, P. S., & Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of the 33rd International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 48, pp. 2139–2148). PMLR. https://proceedings.mlr.press/v48/thomasa16.html

Thomas, P., Theocharous, G., & Ghavamzadeh, M. (2015). High confidence policy improvement. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 37, pp. 2380–2388). PMLR. https://proceedings.mlr.press/v37/thomas15.html

Trendov, N. M., Varas, S., & Zeng, M. (2019). *Digitalisation in agriculture: A scoping review*. Food and Agriculture Organization of the United Nations (FAO). https://www.fao.org/3/ca4985en/ca4985en.pdf

Trevisan, R. G., Martin, N. F., Fonteyne, S., Verhulst, N., Dorado Betancourt, H. A., Jimenez, D., & Gardeazabal, A. (2022). Multiyear maize management dataset collected in Chiapas, Mexico. *Data in Brief, 40*, 107837. https://doi.org/10.1016/j.dib.2022.107837

Vallepogu, N. P., Chaitanya, J., Rani, B. S., & Rohan, R. (2024). Review: A machine learning–based fertilizer recommendation system for sustainable crop yield. *International Research Journal of Education and Technology, 6*(12), 1777–1783.

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., & Gabry, J. (2024). Pareto smoothed importance sampling. *Journal of Machine Learning Research, 25*, 1–58. https://jmlr.org/papers/volume25/19-556/19-556.pdf

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications, 24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

Verhulst, N., Trevisan, R. G., Martin, N. F., Fonteyne, S., Dorado Betancourt, H. A., Jimenez, D., & Gardeazabal, A. (2021). *Agronomic, soil and weather data from modules and extension areas in hub Chiapas, Mexico 2012–2018 (Version 3) [Dataset]*. CIMMYT Research Data & Software Repository Network. https://hdl.handle.net/11529/10548624

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Wang, Y.-X., Agarwal, A., & Dudík, M. (2017). Optimal and adaptive off-policy evaluation in contextual bandits. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 70, pp. 3589–3597). PMLR. https://proceedings.mlr.press/v70/wang17a.html

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Yasabu, S. (2021, September 9). Understanding decision support. CIMMYT. https://www.cimmyt.org/news/understanding-decision-support/

Yu, T., Thomas, G., Yu, L., Ermon, S., Zou, J., Levine, S., Finn, C., & Ma, T. (2020). MOPO: Model-based offline policy optimization. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*. Curran Associates, Inc.

Zhang, R., Zhang, X., Ni, C., & Wang, M. (2022). Off-policy fitted Q-evaluation with differentiable function approximators: Z-estimation and inference theory. In *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)* (Proceedings of Machine Learning Research, Vol. 162, pp. 26713–26749). PMLR. https://proceedings.mlr.press/v162/zhang22al.html

ZOZO Technologies, Inc. (n.d.). *Estimators*. OBP Documentation. Retrieved October 13, 2025, from https://zr-obp.readthedocs.io/en/latest/estimators.html

Appendix A

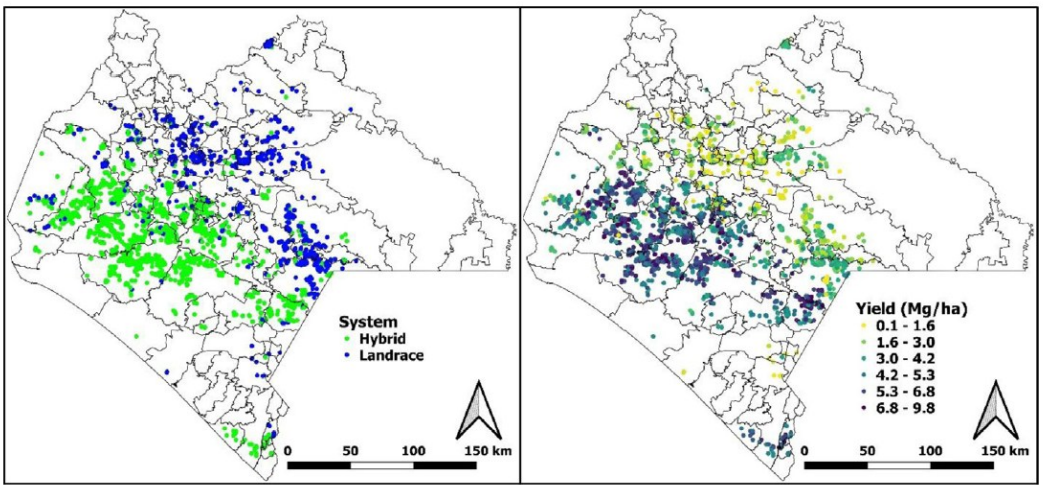

Figure A-1. Spatial distribution of system type and maze yield in Chiapas, Mexico (Trevisan et al., 2022)

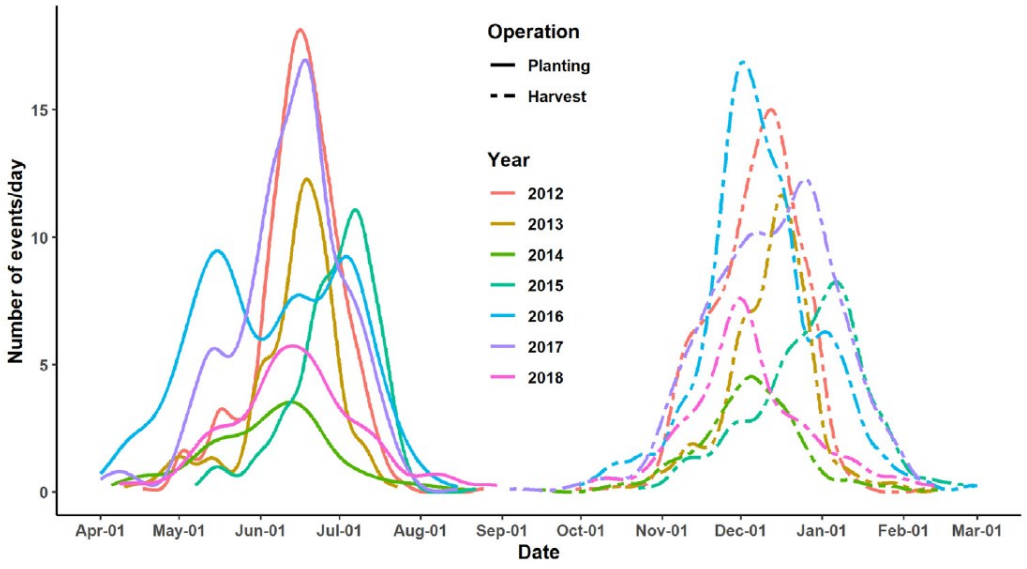

Figure A-2. Temporal distribution of planting and harvesting maize cropping events in Chiapas, Mexico (Trevisan et al., 2022)

Table A-1. Offline evaluation of all policies (MXN/ha)

Estimator	ε	λ	$V_P(\pi_1)$	$V_P(\pi_0)$	Profit Gain Δ (95% CI)	$\hat{k}$
DR	0.0	0.0	10,915	10,236	+679 (+517 to +843)	0.78*
DR	**0.1**	**0.0**	**10,857**	**10,238**	**+619 (+455 to +772)**	**0.63**
DR	0.1	0.1	10,793	10,238	+555 (+389 to +705)	0.63
DR	0.2	0.0	10,797	10,238	+559 (+394 to +709)	0.46
DR	0.2	0.1	10,739	10,238	+501 (+335 to +650)	0.53
SNDR	0.0	0.0	10,884	10,236	+648 (+434 to +852)	0.78*
SNDR	**0.1**	**0.0**	**10,820**	**10,238**	**+582 (+407 to +776)**	**0.63**
SNDR	0.1	0.1	10,771	10,238	+532 (+379 to +705)	0.63
SNDR	0.2	0.0	10,767	10,238	+529 (+357 to +734)	0.46
SNDR	0.2	0.1	10,713	10,238	+475 (+322 to +653)	0.53

*Note: * indicates failed result due to $\hat{k} > 0.7$. Bold rows highlight the primary results with $\varepsilon = 0.1$ and $\lambda = 0.0$. DR = Doubly Robust estimator; SNDR = Self-Normalized Doubly Robust estimator.*