A Causal Machine Learning Dose–Response Framework for Stage-Specific Water–Yield Response and Irrigation Decision Support in Potato Production

Pei-Yu Wu

A thesis

submitted to the faculty of graduate studies

in partial fulfillment of the requirement

for the degree of

Master of Arts

Graduation Program in Information Systems & Technology

York University

Toronto, Ontario

July, 2026

© Pei-Yu Wu 2026

Abstract

Increasing climate extremes create a need for agricultural decision-support tools that move beyond yield prediction toward estimating crop responses to management interventions. This thesis develops a causal machine learning framework to estimate continuous dose–response relationships between stage-specific water deficit/excess and potato yield in Prince Edward Island, Canada, using postal-code-level data from 1982–2024.

The framework integrates causal identification, Generalized Propensity Score diagnostics, common-support trimming, and stage-specific S-learner models to estimate empirically supported dose–response curves and marginal yield responses. Results show strong stage dependence. Mid-season exhibits the clearest response, with the highest marginal returns occurring as severe deficits are relieved between -150 and -100 mm. Early-season sensitivity is moderate, whereas late-season benefits are weaker and water excess is associated with yield penalties. Spatial risk–return analysis further reveals substantial regional heterogeneity. Overall, the findings support stage-specific, soil-informed, and spatially targeted supplemental irrigation, with mid-season deficit relief representing the highest-leverage intervention.

Acknowledgements

I would like to express my heartfelt gratitude to all those who have inspired me along this extraordinary journey. To Dr. Stephen Chen, for welcoming me to his team and begin this work. His constant curiosity and willingness to explore every possible direction have taught me not only how to approach research, but also how to remain open and courageous in life.

I am especially grateful to Dr. Dania Tamayo-Vera, who has always helped me find my way whenever I felt lost. Her energy, clarity, and passion have been a source of inspiration throughout this process. I will always remember our conversation at Strathgartney Provincial Park, where we discussed research methods while hiking through the forest. The fresh air, the movement of the trail, and the exchange of ideas created a kind of spark that reminded me how joyful learning can be.

I would also like to thank Dr. Morteza Mesbah for his patience, generosity, and careful attention to detail. I truly enjoyed every brainstorming session with him, and I am grateful for the time and thought he devoted to helping me refine this work.

To my family, thank you for loving me and supporting me unconditionally. Without your encouragement, I would not have had the courage to challenge myself and continue moving forward. Finally, to Aries, for staying by my side with care and strength throughout this journey.

Contents

Abstract ii

Acknowledgements iii

Table of Contents . . . . . **iv**

List of Figures . . . . . **vii**

List of Tables . . . . . **xii**

Chapter 1: Introduction **1**

1.1 Research Objectives and Questions 4

1.2 Research Gaps and Thesis Contributions 5

1.3 Thesis Structure 7

Chapter 2: Literature Review **8**

2.1 What is Causal Machine Learning? 8

2.2 Applications of Causal ML in Environment and Agriculture 11

2.3 Methodological Challenges and Best Practices in Applying Causal ML to Agriculture 14

2.3.1 Confounding and the Structure of Observational Agricultural Data . . . 15

2.3.2 Continuous Treatments and the Dose-Response Estimation Gap 16

2.3.3 Overlap Diagnostics, Causal Identification, and Refutation 18

2.4 Toward Reproducible, Interpretable, and Generalizable Causal ML for Agriculture 20

2.5 Summary 21

Chapter 3: Materials and Methods 23

3.1 Study Area and Dataset 23

3.2 Stage-specific Variable Decomposition and Feature Construction 27

3.2.1 Stage-specific Decomposition: Start-, Mid-, and Late-season 28

3.2.2 Crop Water Deficit/Excess Estimation 28

3.2.3 Yield Persistence Encoding and Pre-planting Climate Controls 30

3.3 Causal Design and Adjustment Set 32

3.3.1 Positivity Diagnostics and Common Support 35

3.4 S-learner Estimation and Model Validation 39

3.5 Marginal Yield Response and Effective Water-allocation Ranges 41

Chapter 4: Results 43

4.1 Stage-specific Dose–response Curves and Placebo Refutation 43

4.2 Marginal Yield Response and Effective Water-allocation Ranges 50

Chapter 5: Discussion 52

5.1 Implications for Supplemental Irrigation Allocation and Optimization 53

5.2 Model Explainability and the Distinction between Prediction and Intervention . 58

5.3 Soil-related Heterogeneity in Mid-season Response Curves 60

5.4 Application Guidelines for Decision Support 63

5.5 Limitations and Future Directions 64

Chapter 6: Conclusion 69

References 76

Appendix A **84**

Variables used in the causal water–yield response framework. 84

List of Figures

1 **Schematic of the proposed causal machine learning pipeline.** The workflow transitions through five distinct phases: (1) data collection and feature construction; (2) causal identification; (3) positivity assumption violation diagnosis and handling; (4) S-learner model evaluation and dose–response validation; and (5) dose-response function estimation and robustness check. 24

2 **Study area across Prince Edward Island (PEI).** Spatial delineation of the 18 postal-code units, aggregated into four geographic zones for comparative analysis. . 25

3 **Annual potato yield by postal-code grouped region in PEI.** Boxplots illustrating inter-annual dataset variability (1982–2024) of historical marketable potato yields (t/ac). Regional sample sizes are shown below the yield boxplots as the number of postal-code–year observations and the number of contributing postal codes. Boxes denote the interquartile range (IQR), internal lines represent the median, and whiskers extend to 1.5×IQR. 26

4 **Stage-specific water deficit/excess by postal-code grouped region in PEI.** Boxplots illustrating inter-annual dataset variability (1982–2024) of stage-specific water deficit/excess, highlighting consistent mid-season deficits. Boxes denote the interquartile range (IQR), internal lines represent the median, and whiskers extend to 1.5×IQR. 30

5 **Final directed acyclic graph used to identify the causal effect of stage-specific water deficit/excess on potato yield.** The orange nodes represent the treatment and outcome. Green nodes and arrows represent treatment–outcome confounders: pre-planting precipitation and, where applicable, prior-stage water deficit/excess, each of which influences both the current-stage treatment and potato yield. The red arrow represents the causal pathway from stage-specific water deficit/excess to yield. Blue nodes and arrows represent pre-treatment prognostic covariates that directly predict yield but do not cause the current-stage treatment, including historical yield, soil texture, cation-exchange capacity, and soil water-retention characteristics. Prior-stage water deficit/excess is not applicable to the first analyzed growth stage. . 34

6 **GPS distributions by growing-season water deficit/excess decile before and after trimming.** Panel (A) shows the generalized propensity score (GPS) distributions for the full sample before trimming, grouped by realized growing-season water deficit/excess decile. Panel (B) shows the corresponding GPS distributions after GPS-based trimming and refitting the treatment model on the retained sample. Boxes represent the interquartile range, orange lines indicate medians, and whiskers show the non-outlying range. Lower GPS values indicate weaker conditional support for the observed treatment level given the covariate profile. 37

7 **Regional early-season dose–response curves and marginal returns.** Panel A shows regional aggregated dose–response curves for early-season water deficit/excess. The vertical dashed line denotes balanced water availability, where deficit/excess equals zero. Panel B shows the corresponding marginal return across the treatment grid; red indicates positive yield return from additional water availability, blue indicates negative return, and near-white values indicate limited marginal response. 45

8 **Regional mid-season dose–response curves and marginal returns.** Panel A shows regional aggregated dose–response curves for mid-season water deficit/excess. Panel B shows local marginal returns across the same treatment range. Strong positive marginal returns are concentrated under severe-to-moderate deficit conditions, indicating the water-balance interval where additional water availability provides the greatest expected yield gain. 47

9 **Regional late-season dose–response curves and marginal returns.** Panel A shows regional aggregated dose–response curves for late-season water deficit/excess. Panel B shows the corresponding marginal-return heatmap. Positive marginal returns are weaker than in the mid-season case, while negative returns become more evident under late-season excess-water conditions. 48

10 **Placebo refutation based on observed and placebo-shuffled-treatment response spans.** Response span is defined as the vertical range of an estimated dose-response curve. Blue boxes summarize postal-code response spans from the real treatment models, while yellow boxes summarize placebo treatment spans generated by shuffling water deficit/excess values within postal codes, thereby preserving postal-code structure while disrupting the original treatment–yield alignment. Numbers inside boxes denote median response span. Larger real-versus-placebo separation indicates stronger evidence that the estimated curve reflects treatment related signal rather than random model structure. 49

11 **Postal-code-level spatial distribution of mid-season marginal yield return and severe water-deficit exposure across PEI.** Panel A shows the average marginal yield return within the effective mid-season deficit-relief range from -150 to -100 mm, as derived from the estimated postal-code-level dose–response curves; greener shading indicates a larger estimated yield return per additional millimetre of water. Panel B shows the historical exposure frequency of severe mid-season water deficit, calculated as the percentage of postal-code–year observations with water deficit/excess below -100 mm; darker red shading indicates more frequent exposure. 54

12 **Postal-code-level mid-season risk–return irrigation decision plot.** Each point represents one postal code, with colors indicating regional grouping. The x-axis shows the observed frequency of mid-season water deficit below -100 mm, representing historical exposure to severe mid-season deficit. The y-axis shows the average marginal return within the effective deficit-relief range from -150 to -100 mm, representing the expected yield response to additional water availability. Dashed reference lines denote the decision thresholds used in this study: 25% exposure frequency and 0.025 $\mathrm{t/ac/mm}$ average marginal return. 56

13 **SHAP summary for the mid-season S-learner fitted on the trimmed training set.** Each point represents one observation. The x-axis shows the SHAP value, which indicates the contribution of a feature to the model prediction relative to the model baseline. Positive SHAP values increase predicted yield, while negative SHAP values decrease predicted yield. Features are ranked on the y-axis by mean absolute SHAP value. Point color represents the feature value, with blue indicating low values and red indicating high values. The figure should be interpreted as a model-explainability diagnostic rather than as causal evidence; causal interpretation comes from the identified dose–response framework. 59

14 **Associations between soil properties and mid-season response-curve indicators.** Cells report Spearman correlation coefficients between soil variables and indicators extracted from postal-code mid-season dose–response curves. The three curve-derived indicators are effective range width, average marginal return within the effective range, and maximum predicted yield. Red indicates positive association and blue indicates negative association. Significance symbols denote the reported correlation test levels: ** $p < 0.01$, * $p < 0.05$, and † $p < 0.10$. These associations are exploratory and should not be interpreted as causal soil effects. 61

List of Tables

1 Summary of Differences Between Predictive ML and Causal ML 9

2 Regional summary of dominant soil properties (0–60 cm depth). Values represent area-weighted postal-code means within each defined geographic region. 27

3 Held-out predictive performance of candidate S-learner base models for potato yield prediction. 40

4 Summary of stage-specific water-yield response characteristics. 48

5 Summary of yield, climate, and soil variables. 85

Chapter 1: Introduction

Ongoing global warming is causing increasingly erratic climate patterns, while extreme weather events, including heatwaves and droughts, are becoming more frequent and severe [1]. Because these climatic extremes threaten both global food security and national agricultural economies [2], there is an urgent need for agricultural research that can improve crop-yield modeling and support resilient production under increasingly uncertain climatic conditions. In Prince Edward Island (PEI), Canada, potato production accounts for a substantial share of agricultural activity and contributes significantly to the regional economy [3, 4]. Although potato production in the humid climate of Atlantic Canada has traditionally relied primarily on rainfall, the increasing frequency of drought has created growing production risks. Water stress is a major threat to potato production because it can restrict tuber expansion and substantially reduce both yield and quality [5]. Consequently, PEI growers are increasingly adopting supplemental irrigation to mitigate the effects of abnormally dry conditions. Under these circumstances, an irrigation decision-support framework that provides guidance on when irrigation is needed, where limited water resources should be allocated, and how much water should be applied is essential for improving water-use efficiency and maintaining stable yields.

To optimize irrigation management under projected global-warming scenarios [1], artificial intelligence (AI) is increasingly being used to support agricultural decision-making. Machine-learning (ML) methods can achieve strong predictive performance in crop-yield modeling by integrating climate, soil, and remote-sensing information [6–9]. However, predictive accuracy alone is insufficient for informing agricultural management decisions. Standard ML models

trained on observational data primarily capture statistical associations [10–12], and they may therefore conflate the effect of water availability with the influence of confounding factors such as historical rainfall patterns, soil texture, drainage conditions, and management history [13]. Because of this entanglement, associational models cannot reliably answer the interventional questions that are most relevant to producers: if a specific amount of supplemental water is applied during a particular growth stage, what yield gain can be expected, and how might that return vary across locations with different soil and environmental conditions?

Providing actionable irrigation guidance therefore requires a shift from predictive pattern recognition toward causal *what-if* reasoning [10, 14, 15]. Because conventional ML methods do not independently distinguish causal effects from statistical associations, causal-inference principles must be integrated into the modeling workflow. Causal inference has been applied in agriculture to evaluate discrete or binary interventions, such as whether the adoption of a precision-agriculture practice produces measurable productivity gains [14, 16]. Irrigation optimization, however, requires evaluating water availability as a continuous exposure rather than as a simple treated-versus-untreated condition. Characterizing how yield changes across this continuous spectrum therefore requires estimation of a causal dose–response function.

This thesis addresses this research gap by framing potato-yield response to water availability as a continuous dose–response problem. Water availability is represented by the degree of water deficit or excess during the growing season, while yield responses are estimated under alternative levels of this continuous exposure. This framework makes it possible to investigate whether productive ranges of water availability can be identified, whether yield penalties emerge under severe water deficit or excess, whether water sensitivity differs across crop growth stages, and whether these relationships vary spatially across PEI. The empirical analysis draws on a long-term dataset of potato yields from 18 PEI postal codes spanning 1982 to 2024, within a region where potato-yield sensitivity to water availability has been well documented [17, 18].

Estimating causal dose–response relationships from observational agricultural data is challenging for three principal reasons. First, water availability is not randomly assigned; rather, it is determined by weather conditions and interacts with landscape characteristics, soil water-retention capacity, drainage, and management practices, all of which may also affect yield [19,20]. Second, agricultural confounding is high-dimensional because climate indicators, soil properties, geographic location, and historical yield patterns are often correlated and may jointly influence both water exposure and yield outcomes. Third, continuous treatments are particularly susceptible to violations of the positivity assumption. Extreme water deficits or surpluses may occur only under specific combinations of environmental conditions, meaning that certain treatment levels are weakly represented or entirely absent for particular field profiles [10, 14, 21]. Without explicitly diagnosing such support limitations, flexible ML models may extrapolate into poorly represented treatment regions and generate smooth dose–response curves that appear plausible but lack sufficient empirical support.

Recent advances in causal machine learning provide tools for addressing these challenges by combining explicit causal adjustment with flexible nonlinear learning algorithms [10, 12, 13]. These methods make it possible to estimate how outcomes may change under hypothetical interventions while adjusting for observed confounding factors. Accordingly, causal machine learning is used in this thesis not merely to improve potato-yield prediction, but to construct an irrigation decision-support framework. The proposed framework integrates domain-informed causal identification, generalized propensity-score diagnostics, common-support trimming, S-learner dose–response estimation, growth-stage-specific marginal-return analysis, and placebo refutation. Together, these components are intended to move the analysis beyond predictive association and toward empirically supported counterfactual reasoning for water management in potato production.

1.1 Research Objectives and Questions

This thesis investigates the following core research question: How can a causal machine learning dose-response framework estimate spatially and seasonally heterogeneous water-yield relationships in agriculture, and translate these relationships into credible decision support for irrigation scheduling and broader water-resource planning?

Addressing this question requires decomposing the problem into four interrelated sub-questions spanning causal identification, temporal sensitivity, spatial heterogeneity, and practical decision support:

1. **RQ1 — Methodology and Causal Identification:** How can a domain-driven Directed Acyclic Graph (DAG), adjustment-set selection, overlap diagnostics, and robustness checks be used to isolate the causal effect of water deficit and excess from high-dimensional confounding in observational agricultural data?
2. **RQ2 — Phenological Sensitivity:** How do causal yield responses differ when water deficit and excess are evaluated across distinct crop growth stages, and what do these differences imply for irrigation timing decisions?
3. **RQ3 — Spatial Heterogeneity:** To what extent are localized soil properties associated with spatial heterogeneity in water–yield response curves, including effective water ranges, marginal returns, yield plateaus, and downside risks under extreme water conditions across PEI postal codes?
4. **RQ4 — Actionable Decision-Making:** How can nonlinear causal dose–response curves be translated into a practical decision layer that identifies stage-specific, and region-specific water-allocation priorities?

1.2 Research Gaps and Thesis Contributions

The literature reviewed in Chapter 2 exposes two structural gaps that jointly motivate this thesis. The first is a methodological gap. Agricultural studies that model water availability as a continuous input-most notably Jiang et al.'s [17] polynomial dose–response analysis of potato yield in Atlantic Canada, the study estimate graded water–yield relationships but do not include a causal identification step. As a result, confounding by soil type, management history, spatial heterogeneity, and persistent yield differences remains insufficiently addressed. Conversely, agricultural causal ML studies that do prioritize identification, including evaluations of recommendation systems [16], soil health assessments [22], and yield modeling under distributional shift [11], typically frame interventions as binary or categorical. This leaves the graded, continuous structure of many agronomic inputs outside the causal framework.

The second gap concerns workflow integration. The technical toolkit for continuous-treatment causal inference—including generalized propensity score (GPS) modeling, average dose–response function (ADRF) estimation, overlap diagnostics, common-support treatment range restriction, and placebo validation—has been developed and applied primarily in medical, epidemiological, and econometric research [23–26]. However, these tools have not yet been translated into an end-to-end, reproducible agricultural causal workflow for estimating continuous water–yield response under realistic observational data constraints.

This thesis responds directly to both gaps by developing and applying a complete continuous-treatment causal workflow for potato yield response to water deficit and excess. The contribution is not limited to fitting a flexible yield model. Rather, the thesis integrates causal identification, overlap assessment, dose–response estimation, stage-specific marginal-return analysis, spatial heterogeneity interpretation, and decision-support translation into a single framework. Specifically, the thesis makes four interrelated contributions, each corresponding to one of the research

questions.

In response to **RQ1**, the thesis bridges the causal identification gap by constructing a domain-driven DAG to define a defensible adjustment set for the water availability–yield relationship in PEI potato production. Rather than selecting covariates only by predictive importance, the framework uses causal structure and the backdoor criterion to distinguish confounders from mediators, colliders, and outcome-only predictors that could bias the estimated effect if handled improperly. GPS-based overlap diagnostics operationalize the positivity requirement by detecting and trimming regions of the treatment–covariate space with insufficient empirical support. This ensures that the reported dose–response curves are estimated over supported treatment ranges rather than being driven by extrapolation. Placebo refutation tests provide an additional robustness check by evaluating whether the estimated curves retain meaningful treatment-related signal after comparison with shuffled-treatment models.

In response to **RQ2**, the thesis extends dose–response function estimation into three discrete phenological windows: early-season, mid-season, and late-season. This stage-specific decomposition allows the analysis to determine whether water deficit and excess have different yield implications depending on when they occur. Instead of estimating a single growing-season water effect, the thesis evaluates whether yield sensitivity is concentrated in specific periods of crop development. This provides irrigation-timing insight with direct operational relevance.

In response to **RQ3**, the thesis treats spatial heterogeneity as a core inferential target rather than as noise to be averaged away. By examining how localized soil properties are associated with the response heterogeneity indicators—including effective water range width, average marginal return within the effective range, and maximum predicted yield—the thesis connects regional water-response patterns to soil context. This makes it possible to interpret not only whether water matters, but also where the water–yield relationship is stronger, weaker, broader, or more vulnerable to excess moisture.

In response to **RQ4**, the thesis closes the workflow by translating estimated dose–response curves into a practical decision-support perspective. Marginal return estimates are used to identify effective water-allocation ranges, while regional risk–return analysis compares locations according to both deficit exposure and expected yield gain from deficit relief. This translation step connects causal estimates to irrigation planning by identifying where supplemental water is most likely to provide agronomically meaningful returns. The resulting framework provides a foundation for future irrigation optimization models that incorporate water availability, infrastructure constraints, pumping cost, and economic return.

1.3 Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2 presents the systematic literature review, establishing why a continuous-treatment causal workflow is methodologically necessary yet currently underdeveloped in agriculture. Chapter 3 describes the study area, data pipeline, DAG-based causal identification, GPS-based positivity trimming, S-learner estimation, and placebo validation procedures, firmly addressing RQ1. Chapter 4 presents the empirical results, comparing stage-specific dose–response curves and marginal-return heatmaps to address RQ2. Chapter 5 discusses the implications for spatial heterogeneity and decision support, examining soil-related variations and translating the marginal-response functions into regional irrigation-prioritization strategies, addressing RQ3 and RQ4. Finally, Chapter 6 revisits the research questions, summarizes the main contributions, discusses limitations, and identifies future directions for field-level causal estimation and climate-adaptive agriculture.

Chapter 2: Literature Review

As digital agriculture increasingly relies on remote sensing, IoT sensors, farm records, and climate data, machine learning has become a powerful tool for forecasting and pattern detection. Yet agricultural decision-making often requires more than accurate prediction alone. Researchers and practitioners increasingly need to understand not only what is likely to happen, but also why outcomes occur, which factors are actionable, and how outcomes may change under alternative management or environmental conditions. In this context, the literature reviewed in this chapter explores the conceptual foundations of Causal Machine Learning (Causal ML), its applications in environmental and agricultural research, the methodological challenges involved in applying it to observational agricultural data, and the emerging need for more reproducible, interpretable, and generalizable causal workflows.

2.1 What is Causal Machine Learning?

The growing use of digital agriculture technologies has transformed machine learning from a niche analytical tool into a central component of modern agricultural research. With increasing access to remote sensing, IoT sensors, farm records, and climate data, machine learning models are now widely used to predict yield, detect crop stress, classify land use, and support management decisions [6–8]. This expansion has improved researchers' ability to extract patterns from complex agricultural systems, but it has also exposed an important limitation of predictive modeling: accurate prediction does not necessarily reveal why an outcome occurs, nor does it indicate how that outcome can be changed through management or policy intervention.

This limitation has motivated increasing interest in Causal ML. Broadly, Causal ML refers to machine learning approaches that incorporate causal inference principles to model how data are generated and how outcomes respond when a system is deliberately changed. Rather than learning associations from observed data alone, these methods use causal frameworks such as Structural Causal Models (SCMs) to distinguish genuine cause-and-effect relationships from correlations that arise incidentally from the data [10]. In this way, Causal ML extends machine learning beyond pattern recognition toward explanation, intervention, and counterfactual reasoning [13]. It is therefore intended not only to answer what is likely to happen, but also what would happen if conditions were changed, and whether a different action would have produced a different result (Table 1).

Feature	**Predictive ML**	**Causal ML**
Statistical Basis	Correlation and Association [11, 12]	Causality and Mechanism [10, 11]
Primary Question	What is Y given I see X? [10]	What is Y if I do X? [10, 16]
Robustness	Sensitive to distribution shifts [10, 12]	Robust through invariant features [13, 21]
Interpretability	Often a black box [8, 15]	Inherently more explainable [11, 12]
Data Reliance	Passively collected data [10]	Interventional or deconfounded observational data [10, 13]

Table 1: Summary of Differences Between Predictive ML and Causal ML

The distinction between predictive ML and Causal ML is especially important because the two paradigms address fundamentally different questions. Conventional ML is primarily associational—it identifies statistical regularities in historical observations and performs best when future data resemble the training environment. This makes it highly useful for forecasting, but also vulnerable to spurious correlations and unstable relationships across regions, seasons, and production contexts [12]. Causal ML, by contrast, is interventional in orientation. Its goal is

to isolate the effect of a change in one variable on another while accounting for confounding structure, thereby identifying relationships that are more meaningful for decision-making. A predictive model may indicate that yield will decline under certain conditions, but a causal model is better suited to estimate whether changing irrigation, sowing date, fertilizer use, or crop rotation would alter that outcome and by how much [16].

This causal perspective is particularly important in agriculture because agriculture is not only a forecasting domain, but also a decision domain. Agroecosystems are shaped by interacting climatic, soil, biological, and management processes that vary across both space and time, making purely associational models difficult to generalize across environments [13]. At the same time, many agricultural questions are inherently interventional: farmers and policymakers need to know how outcomes respond when management practices, input levels, or environmental conditions change. Prior studies therefore position Causal ML not as a replacement for machine learning, but as a necessary extension of it. As agricultural research moves from prediction toward decision support, the central issue is no longer only whether a model can predict outcomes accurately, but whether it can identify stable and actionable relationships that remain credible under changing conditions [16]. This reorientation is increasingly framed in the literature under the concept of Trustworthy AI—a multi-dimensional standard encompassing stability across environments, explainability of model decisions, robustness under distributional shift, and fairness in the treatment of heterogeneous subgroups [10, 12, 15]. Causal ML is positioned as a primary technical path toward this standard, because grounding model behavior inv a causal mechanism rather than correlation enables formal validation of stability, interpretability, and robustness in ways that purely associational approaches cannot support [12, 13]. The literature reviewed in this chapter is therefore read not only as a survey of methods, but as a cumulative argument for why trustworthy agricultural AI requires causal grounding.

2.2 Applications of Causal ML in Environment and Agriculture

The application of Causal ML in environmental and agricultural research has developed in response to a shared challenge: conventional machine learning can identify patterns in complex natural systems, but it cannot reliably separate physical drivers from confounded associations, nor can it estimate how outcomes would change under deliberate intervention. The literature has addressed this challenge progressively, first by recovering causal structure in environmental systems where feedback and non-stationarity distort simple correlations, then by embedding structural knowledge into agricultural models designed to support intervention and resource allocation. Across this progression, the unifying contribution of Causal ML is not improved predictive accuracy alone, but the ability to generate mechanistic, transferable, and decision-relevant knowledge in systems that are nonlinear, dynamic, and highly context-dependent.

One important strand of this work addresses the fundamental difficulty of attributing causation in coupled land–atmosphere and hydrological systems, where observed associations are routinely distorted by seasonal cycles, spatial autocorrelation, and bidirectional feedbacks. To overcome this, nonlinear Granger causality and Random Forest-based causal inference have been applied to estimate the effect of soil moisture on rainfall occurrence while controlling for climate memory [27]. Similarly, time-series causal discovery methods such as PCMCI (Peter and Clark Momentary Conditional Independence) have been used to identify lagged causal links among meteorological, soil moisture, and agricultural drought indicators [28]. In these environmental applications, the contribution of Causal ML is not improved forecast accuracy, but the recovery of temporal and physical structure, a distinction that is essential when the goal is to understand what drives a system rather than merely to extrapolate its future state.

Building on this, causal ML has been extended into crop modeling to address a specific robustness failure of purely data driven approaches: their tendency to break down under conditions outside

the training distribution, including the extreme weather events that are increasingly common under climate change. In soybean yield estimation, Structural Causal Models have been combined with Graph Attention Networks in the SCM-GAT framework to encode causal relationships among weather variables, vegetation indices, and yield formation processes [11]. By anchoring a flexible deep learning architecture to structural causal knowledge, the model achieved greater stability under extreme conditions than purely associational alternatives. This illustrates a broader principle in the literature: causal ML is most powerful not in opposition to machine learning, but as a structural constraint that prevents flexible models from learning brittle, context-specific patterns that do not transfer across environments.

While recovering physical structure is vital for explanation, a distinct challenge arises when the objective shifts to operational decision support, requiring model outputs to be connected to actual management levers rather than left as unexplained predictions. Studies on irrigation water use forecasting have addressed this by coupling PCMCI with deep learning architectures to improve short-term water demand prediction while filtering out spurious relationships caused by weather autocorrelation [29]. In precision irrigation more broadly, machine learning approaches have been used to identify optimal water application policies [30]. However, a critical limitation runs through this body of work: although irrigation is inherently a continuous variable, and although the agronomic objective is precisely to identify the optimal dose of water across a graded range rather than simply a binary treatment versus control comparison, these studies rely on predictive ML without a causal identification step. The dose-response patterns they produce describe statistical associations between water availability and yield, not the causal effect of intervening on water supply while holding background conditions fixed.

The most explicit use of causal reasoning in agriculture appears in studies that directly evaluate the effects of specific interventions from observational data. Causal inference methods including matching, inverse propensity score weighting, and meta-learners have been applied to assess

whether agricultural recommendation systems genuinely improve farm outcomes, with Tsoumas et al. [16] demonstrating a yield gain of 12% to 17% among cotton farmers who followed sowing recommendations. At a broader scale, established causal designs—Difference-in-Differences, Instrumental Variables, and Synthetic Control—have been used to evaluate policy interventions that cannot be studied through randomized experiments [14], and causal discovery ensembles have been applied to identify interacting socio-environmental drivers of food insecurity [31]. These studies represent the strongest use of causal identification in agricultural ML to date. Yet they share an important structural feature: the interventions they evaluate are framed as binary or categorical, a recommendation either was or was not followed, a policy either was or was not implemented. The continuous, graded nature of most real agricultural inputs, including water availability, fertilization rates, and sowing timing, falls outside the causal framework these studies employ.

More recent work has extended causal analysis toward interpretable assessment of agroecosystem condition, where the objective is not to evaluate a single intervention but to clarify how multiple interacting system components jointly shape long-term outcomes. The CI-SHAP-NA framework combines Double Machine Learning with SHAP values and network analysis to distinguish transient fertilization effects from more intrinsic soil health indicators such as soil organic carbon and available iron [22]. This direction shows that causal analysis in agriculture is increasingly being used to clarify system structure and generate knowledge useful for long-term management—moving beyond isolated binary treatment evaluations toward a more holistic understanding of agroecosystem dynamics. Even so, like the intervention-evaluation studies before it, this work frames inputs as discrete contrasts rather than as graded continuous exposures.

Collectively, the literature shows that Causal ML has been applied across environmental and agricultural research to overcome specific methodological hurdles: recovering physical structure from correlated observations, stabilizing models under distributional shift, and connecting

analytical outputs to management decisions. Across all of these contributions, the shared movement is away from pattern recognition alone and toward models that are more physically grounded, intervention-oriented, and useful for real-world decision-making. At the same time, the survey reveals a consistent structural gap: agricultural causal studies have been designed around binary or categorical treatment comparisons, while agricultural studies that do model continuous dose-response relationships rely on predictive methods that leave confounding unaddressed. The challenge of estimating causal dose-response functions under continuous treatments—which requires specialized identification strategies, generalized propensity score methods, and flexible nonlinear estimators—has not been systematically addressed in the agricultural causal literature. This gap frames the methodological challenge taken up in the following section.

2.3 Methodological Challenges and Best Practices in Applying Causal ML to Agriculture

A recurring theme in the literature is that the methodological difficulty of applying Causal ML in agriculture begins not with the choice of algorithm, but with the nature of the data itself. Most agricultural studies rely on observational rather than randomized data, because treatments such as irrigation, fertilization, or management decisions are rarely assigned at random in real farming systems [10,32]. This creates a fundamental tension—although Causal ML is intended to estimate intervention effects, the available data are generated under real-world decision processes shaped by climate, soil, management history, and farmer behavior. Causal analysis in agriculture must therefore confront not only statistical complexity, but also the fact that treatment assignment is itself embedded within the system being studied.

2.3.1 Confounding and the Structure of Observational Agricultural Data

The first major challenge that follows from this setting is confounding, and its character in agroecosystems is particularly vicious because treatment assignment is structurally entangled with the very factors that drive yield outcomes. Unlike a controlled experiment where water availability can be randomized across plots, observational agricultural data reflects the joint product of climate, soil, and farmer behavior accumulated over years. Fields experiencing greater drought stress may also have lower soil water-holding capacity and shallower root zones, making them inherently less productive—yet these are precisely the fields where more irrigation is likely to be applied. Conversely, historically high-yielding fields often reflect a compounding of favorable soil texture, long-term management intensity, and localized micro-climatic advantages, none of which are fully observable from standard datasets. Unobserved spatial heterogeneity at the sub-field level—including localized differences in drainage capacity, topographic position, and historical accumulation of organic matter—acts as a hidden confounder that simultaneously shapes both the water environment a crop experiences and the yield outcome it produces [12, 14]. In this setting, a naive comparison of yield across water availability levels does not estimate a causal effect; it conflates the impact of water supply with the background differences between locations that determined how much water each received in the first place.

In response, prior studies emphasize that credible causal analysis must begin with explicit identification rather than algorithmic estimation. Directed Acyclic Graphs (DAGs) are used to encode domain knowledge about how treatments, confounders, mediators, and outcomes are related, helping researchers determine which variables must be adjusted and which should be excluded to avoid conditioning on post-treatment variables or colliders [14, 16]. In agricultural systems, where many correlated variables coexist and feature-rich models may otherwise adjust for descendants of the treatment or open non-causal paths, this structural step is particularly

consequential, since incorrect adjustment sets can amplify rather than reduce bias, and the error may not be detectable from predictive performance alone.

Once a valid identification strategy is defined, the estimation step must balance comparability with flexibility. Propensity score methods—matching, inverse probability weighting, and stratification—are used to approximate the balance that randomization would otherwise provide [23, 33–35]. For estimating heterogeneous treatment effects, meta-learners such as the S-learner, T-learner, and X-learner decompose causal estimation into supervised learning tasks, embedding flexible ML models within a valid causal design [16, 36]. However, these standard tools are designed for binary or categorical treatments. The propensity score in its classical form estimates a probability of treatment assignment between two discrete conditions, and meta-learners are typically applied to compare defined treatment groups. When the treatment of interest is continuous—as is the case for water availability, fertilization rates, or growing-degree accumulation—neither the standard propensity score framework nor the conventional meta-learner structure is directly applicable without fundamental modification.

2.3.2 Continuous Treatments and the Dose-Response Estimation Gap

This limitation points to a specific and consequential gap in the agricultural Causal ML literature. Agricultural interventions are inherently graded quantities, and the agronomic objective is typically not simply to identify whether an input helps, but how much is optimal and where that optimum varies across locations and soil conditions. The appropriate causal estimate for such questions is the Average Dose-Response Function (ADRF), which maps every possible treatment level to its expected potential outcome across the target population [24, 25]. Estimating the ADRF from observational data requires the Generalized Propensity Score (GPS)—the conditional density of the continuous treatment given covariates—to characterize and correct for

dosage assignment bias across the full treatment distribution [37, 38]. Specialized architectures such as the Dose-Response Network (DRNet) have been developed to recover individual-level dose-response curves by stratifying the continuous treatment range and learning representations tailored to each stratum [23], and end-to-end balancing approaches have extended these methods to enforce covariate balance across the full continuous treatment distribution rather than at a single threshold [39]. However, these methodological advances have been developed and evaluated almost exclusively in medical and pharmacological contexts.

For any given field in a given season, only one treatment level is ever observed, while the yield under any alternative water availability remains permanently unobserved [10, 23]. This missing-counterfactual problem is especially consequential for continuous treatments, since binary causal questions require recovering one missing potential outcome per unit, continuous dose-response estimation requires recovering an entire function of potential outcomes across the treatment range, the vast majority of which are unobserved for any given location [35]. Unlike medical dose-response problems, which typically concern a single compound administered under controlled clinical conditions, agricultural dose-response relationships are shaped by the non-monotonic interaction of a physical environmental variable with spatially heterogeneous soil and management systems. Too little water causes drought stress; too much causes waterlogging and oxygen stress; and the threshold between these regimes shifts with soil texture, drainage capacity, and root zone depth [17]. This non-monotonicity means that a single flexible estimator trained without structural constraints may fit the observed associations well without recovering the underlying causal curve.

Agricultural studies that model water availability as a continuous variable have not adopted the causal toolkit developed for these problems. Jiang et al. [17] treated growing-season precipitation as a continuous treatment and used polynomial regression to map the full yield dose-response curve for potato in Atlantic Canada, explicitly identifying a sensitive range and an insensitive

plateau. While this study demonstrates a non-linear continuous dose-response reasoning, it relies on regression without a causal identification step, the resulting curve describes a statistical association rather than the causal effect of intervening on water supply. On the other side, agricultural Causal ML studies that prioritize identification, including those that evaluate recommendation systems [16], soil health [22], and yield under extreme weather [11], frame their interventions as binary or categorical, leaving the continuous, graded structure of most agronomic inputs outside the causal framework. The result is a two-sided gap: agricultural dose-response studies are causally uninformed, and agricultural causal studies are structurally binary. Bridging this gap by integrating causal design and dose-response estimators within a principled agricultural causal workflow, constitutes the central methodological challenge that motivates the present thesis.

2.3.3 Overlap Diagnostics, Causal Identification, and Refutation

Whether treatments are binary or continuous, the literature places strong emphasis on diagnosing and repairing weak data support before interpreting estimated effects. Positivity violations are routinely encountered in regional agricultural datasets, since severe drought conditions rarely overlap with certain soil management regimes in historical records, and extreme water surplus may only be observed in specific low-lying, poorly drained areas [21, 40]. As a result, large portions of the joint treatment-covariate space are empirically empty, and any estimator that reaches into these regions is extrapolating beyond the data rather than estimating within it. For continuous treatments, this task is more demanding than in the binary case, since the GPS must remain non-negligible not at a single threshold but across the full treatment range and the full covariate distribution. When positivity is poor, researchers may trim observations with extreme GPS values to enforce common support, restricting inference to a better-supported overlap population [26, 40]. This practice reflects a principle that is especially consequential

for agricultural dose-response analysis and that has direct implications for the nature of model explainability itself. A causal explanation is only as credible as the empirical support behind it: dose-response curves that extend into regions of poor overlap are not causal explanations but statistical extrapolations, and presenting them as actionable guidance overstates what the data can honestly support.

Because the true causal effect is unobservable, prior studies also emphasize the importance of refutation and sensitivity analysis. Placebo treatments, random common causes, and other falsification tests are used to examine whether estimated effects persist when the causal signal should disappear [13, 16]. These procedures do not prove that an estimated effect is correct, but they help assess whether the result is driven by spurious structure, unstable assumptions, or model artifact. In agricultural contexts, where observational data are noisy and causal assumptions are difficult to verify directly, such stress testing is part of the methodological standard for building trust in causal findings.

Taken together, the three methodological challenges documented in this section—confounding rooted in the joint structure of climate, soil, and farmer behavior; missing counterfactuals that multiply across a continuous treatment range; and positivity violations that are spatially and agronomically structured—cannot be resolved by any single estimator. The field has converged on a set of principled practices, but these practices have been developed and applied primarily within binary treatment frameworks. The continuous, graded nature of most agricultural inputs remains largely outside the established causal workflow, constituting a gap that is both methodologically specific and practically consequential.

2.4 Toward Reproducible, Interpretable, and Generalizable Causal ML for Agriculture

Although Causal ML is gaining momentum in climate science and ecology [13], its adoption in agriculture remains limited. This slower uptake stems from the inherent complexity of agroecosystems—which are high-dimensional, nonlinear, and non-stationary—making credible evaluation difficult and the transferability of causal claims uncertain. Furthermore, a tension exists between automated machine learning, which extracts highly predictive but often weak physical meaning features, and the agronomic need for models that support defensible, interpretable reasoning. The challenge is not merely applying causal methods, but deploying them in a reproducible workflow that remains robust across diverse environments.

While the literature has established core principles for credible analysis—structural identification, flexible estimation, and diagnostic validation—few studies assemble these into a cohesive, end-to-end pipeline. Crucially, the methods required to model continuous inputs remain missing from the agricultural causal workflow. As long as continuous agronomic inputs are either modeled without causal identification [17] or artificially discretized [11, 16], the field lacks a procedural template for credible, graded decision support. Without integrating the continuous methodologies proven in econometric and medical research [23, 24], it remains difficult to reproduce findings or verify if estimated dose-response relationships are causally credible beyond their training data.

Closing this structural disconnect defines the future direction of the field. The next generation of agroclimatic modeling must move beyond increasingly accurate black-box predictors toward models built specifically for intervention. In this context, interpretability is not simply post-hoc feature attribution, but the capacity to provide empirically supported answers to counterfactual questions under clear assumptions [12, 21]. Similarly, generalizability extends beyond hold-out

accuracy to the stability of physical relationships across shifting management regimes. Recent calls for hybrid models, explicit uncertainty quantification, and human-centered tools underscore the need to bridge predictive algorithms with field practice [9, 30, 41]. Ultimately, connecting flexible ML with a principled continuous causal framework is essential for translating agricultural data into trustworthy precision management.

2.5 Summary

The literature review has established that estimating causal dose-response functions for continuous agricultural treatments requires a fundamentally different methodological approach than either predictive ML or existing binary-treatment causal frameworks provide. While agriculture is inherently an interventional domain that requires graded, counterfactual reasoning to support farmer decision-making, the current literature remains structurally divided. Studies modeling continuous inputs typically rely on predictive regression, leaving profound spatial and climatic confounding unaddressed. Conversely, rigorous agricultural Causal ML applications have overwhelmingly framed interventions as discrete or binary choices, failing to capture the graded nature of agronomic reality.

Bridging this divide requires recognizing that observational agricultural data is structurally entangled: treatment assignment is driven by the very spatial and climatic heterogeneities that dictate yield. Untangling these relationships requires integrating tools historically confined to medical and econometric research, specifically ADRF estimation and GPS overlap enforcement. The translation of these methods into an agricultural domain characterized by non-monotonic responses and regional confounding represents a critical, unmet methodological need.

These insights directly motivate the design of this thesis. The following chapters propose an end-to-end methodological framework that integrates structural causal identification, GPS-based

overlap trimming, and flexible S-learner estimation into a unified workflow applied to potato yields in PEI.

Chapter 3: Materials and Methods

The thesis proposes a causal machine learning framework to investigate the spatially and seasonally heterogeneous water–yield relationships, and translate these relationships into credible decision support for irrigation scheduling and broader water-resource plannig. As summarized in Fig. 1, the analysis begins by integrating yield records, meteorological observations, and soil attributes into a postal-code–year agroclimate dataset. These records are then transformed into stage-specific water deficit/excess treatments, historical yield controls, and pre-planting climate indicators. The chapter next formalizes the causal design using a directed acyclic graph (DAG) and a backdoor-based adjustment strategy, followed by positivity diagnostics to restrict estimation to the common-support population. Finally, candidate S-learner models are trained and validated using a time-based train–test split, after which the selected estimator is used to generate stage-specific dose–response curves, placebo refutation results, and marginal yield-response summaries, ultimately delineating optimal irrigation targets and spatial heterogeneity across PEI.

3.1 Study Area and Dataset

This study focuses on Prince Edward Island (PEI), Canada's smallest province (5,670 km^2) yet one of its premier potato-producing regions. With 42.5% of its landmass under cultivation, potatoes occupy over 20% of the cropland and represent approximately one-quarter of the national supply [1, 17, 42]. Characterized by a humid temperate climate, PEI exhibits substantial spatial variability in soil properties and growing-season moisture. While historically operated as a rain-fed system, regional yields remain highly vulnerable to fluctuations in both the timing

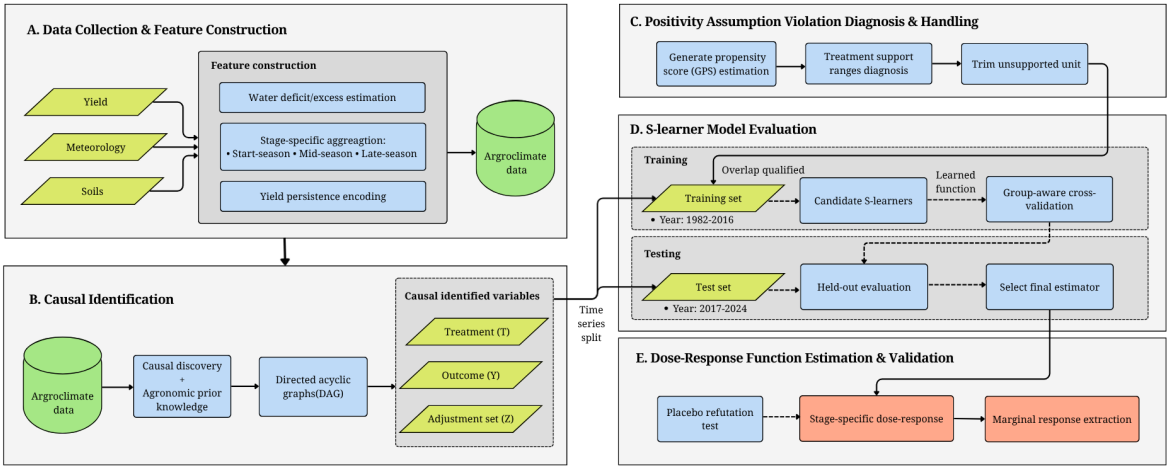

Figure 1: ***Schematic of the proposed causal machine learning pipeline.*** *The workflow transitions through five distinct phases: (1) data collection and feature construction; (2) causal identification; (3) positivity assumption violation diagnosis and handling; (4) S-learner model evaluation and dose–response validation; and (5) dose-response function estimation and robustness check.*

and magnitude of water availability. To capture these dynamics, agroenvironmental data were compiled at the postal-code level. For presentation clarity in spatial comparisons, the 18 constituent postal codes are grouped into four primary geographic regions: northwest (NW), central-west (CW), central-east (CE), and region east, as mapped in Fig. 2.

Annual potato yield records for 1982–2024 were obtained from the PEI Agriculture Insurance Corporation (PEIAIC). The data reported postal-code-level marketable yield in tonnes per acre (t/ac), with each estimate representing at least four farms and aggregating production across commercial varieties such as Russet Burbank, Yukon Gold, and Superior. Each postal-code–year unit links historical productivity to corresponding climate and soil conditions. As shown in the historical distributions (Fig. 2B, bottom), productivity varies notably by geography: Region east consistently performs highest (mean: 13.42 t/ac; IQR: 12.0–14.6 t/ac), whereas Region NW exhibits the lowest baseline productivity (mean: 11.45 t/ac; IQR: 9.9–13.2 t/ac).

Meteorological records were utilized to characterize seasonal crop water demand. Daily measurements of precipitation, air temperature, relative humidity, wind speed, and solar radiation were sourced from Environment and Climate Change Canada (ECCC) ground stations, matching

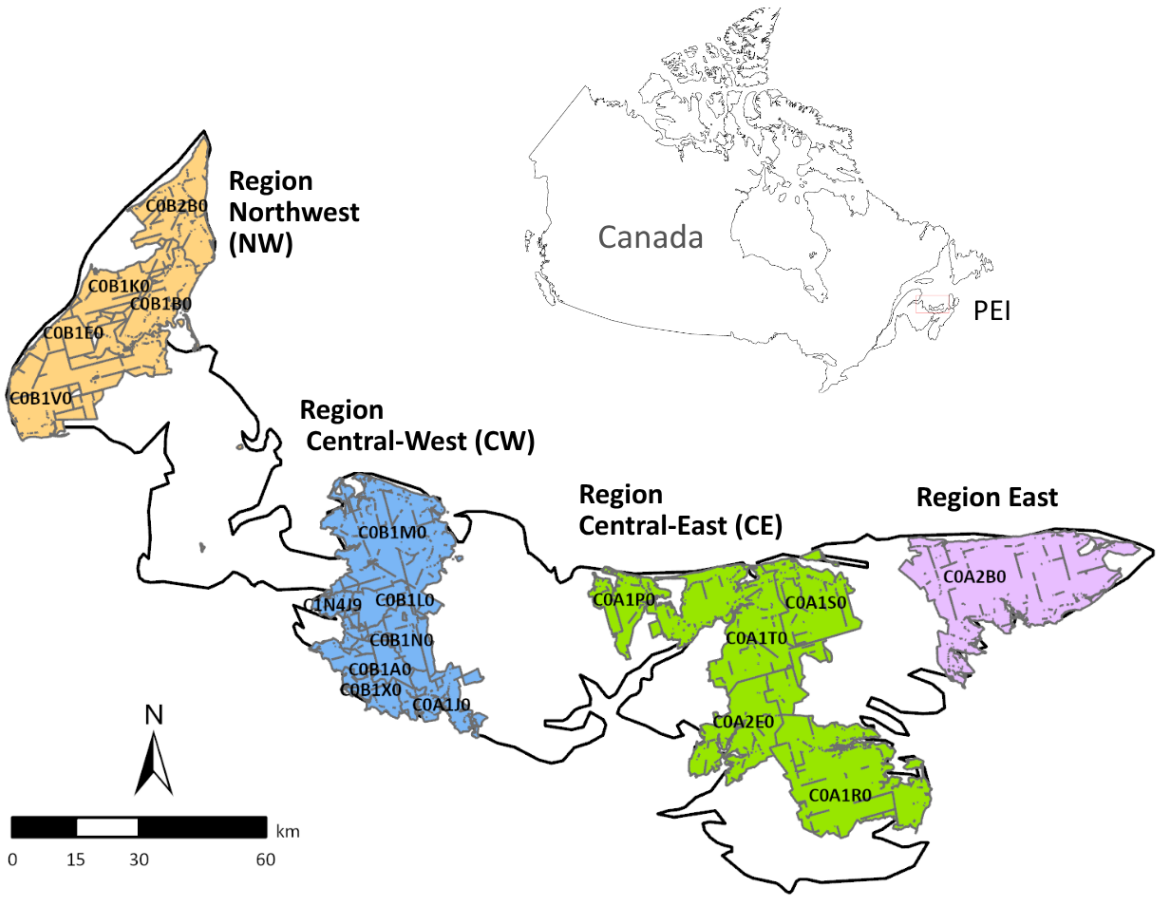

Figure 2: ***Study area across Prince Edward Island (PEI).*** *Spatial delineation of the 18 postal-code units, aggregated into four geographic zones for comparative analysis.*

each postal code to the nearest station within a 30 km radius. Temporal gaps were imputed using NASA Prediction of Worldwide Energy Resources (POWER) data to preserve continuity across PEI's west-east climatic gradient. These daily records were then aggregated over distinct phenological sub-periods to construct the stage-specific water availability indicators described in Section 3.2.

Soil properties were extracted from the Canadian Soil Information Service (CanSIS) [43]. Postal-code-level profiles were derived via area-weighted aggregation, with attributes computed as thickness-weighted horizon means over the 0–60 cm depth interval to isolate the active potato rooting zone [44]. Variables included texture fractions (total sand, very fine sand, total silt, and total clay), coarse fragments, bulk density, cation exchange capacity (CEC), and volumetric

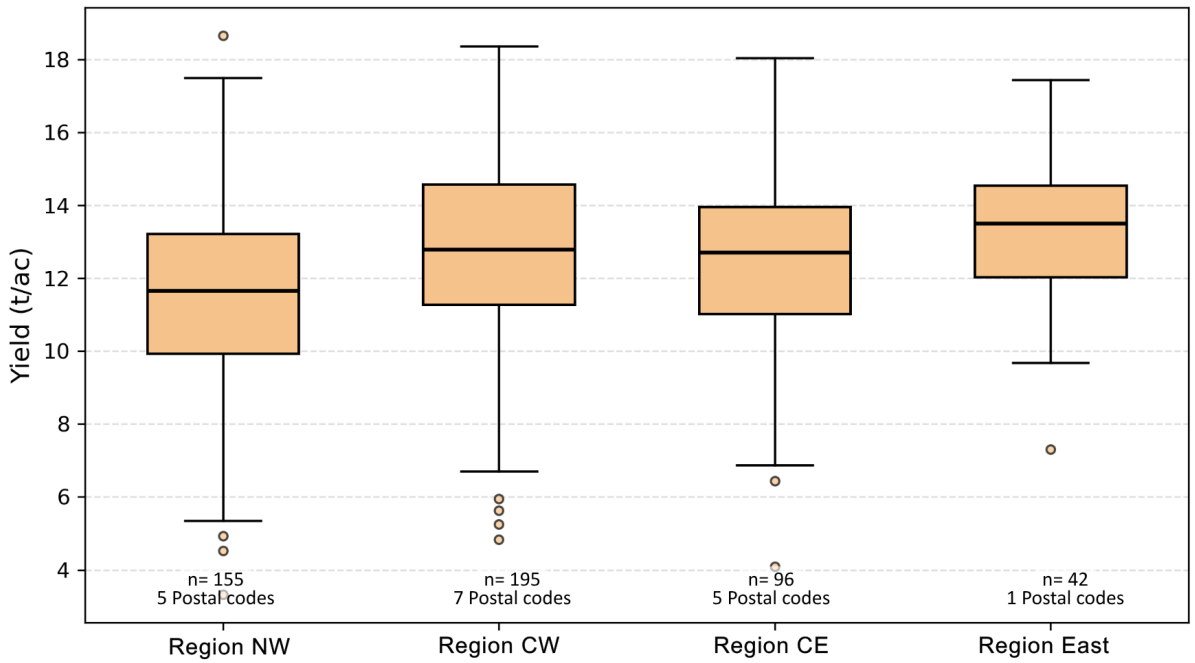

Figure 3: ***Annual potato yield by postal-code grouped region in PEI.*** *Boxplots illustrating inter-annual dataset variability (1982–2024) of historical marketable potato yields (t/ac). Regional sample sizes are shown below the yield boxplots as the number of postal-code–year observations and the number of contributing postal codes. Boxes denote the interquartile range (IQR), internal lines represent the median, and whiskers extend to 1.5×IQR.*

water content at both field capacity and permanent wilting point (Table 2).

By standard pedological definition, the fundamental soil textural class is dictated solely by the fine earth fraction (particles < 2 mm), where the relative proportions of total sand, total silt, and total clay sum to 100%. Within this composition, very fine sand is quantified as a specific functional sub-component of the total sand fraction. Coarse fragments (> 2 mm), in contrast, are evaluated independently as a volumetric proportion of the whole soil matrix. Based on these regional fine-earth profiles, the island's agricultural root zones are predominantly characterized as a sandy loam soil type.

These static features are crucial for modeling spatial heterogeneity, as they represent spatial differences in water retention, drainage capacity, rooting conditions, and organic matter. Specifically, because coarse fragments are rock and gravel particles lacking internal porosity, they physically reduce the active soil volume available for water storage. Meanwhile, the very

fine sand (0.053–0.106 mm) fraction heavily influences micro-porosity within the fine earth matrix, thereby enhancing plant-available water retention under unsaturated conditions [45]. Bulk density is the mass of dry soil per unit volume (g/cm^3); serving as a standard indicator of soil compaction, elevated bulk densities mechanically restrict potato root penetration and tuber expansion, limiting crop growth [46]. Chemically, CEC measures the soil's capacity to retain positively charged nutrient ions (e.g., Ca^{2+}, K^+) on clay and organic matter surfaces, acting as a primary proxy for baseline soil fertility [47]. Finally, soil water dynamics are defined by the interval between field capacity (moisture retained against gravity at a matric potential of -33 kPa) and the permanent wilting point (the threshold where moisture is held too tightly for root extraction, at -1500 kPa). The quantitative difference between these two thresholds dictates the total Plant-Available Water (PAW), which fundamentally drives regional drought resilience and irrigation sensitivity [48].

Table 2: Regional summary of dominant soil properties (0–60 cm depth). Values represent area-weighted postal-code means within each defined geographic region.

Soil Property	**Region NW**	**Region CW**	**Region CE**	**Region east**
Coarse fragments (%)	8.0	13.3	8.6	10.8
Very fine sand (%)	13.3	15.6	17.0	15.9
Total sand (%)	56.5	59.6	65.5	64.1
Total silt (%)	30.6	29.4	24.4	25.7
Total clay (%)	11.7	10.6	8.5	8.7
CEC (meq/100g)	4.4	5.0	5.1	5.5
Bulk density (g/cm^3)	1.4	1.3	1.4	1.3
Field capacity (% v)	43.3	47.4	44.6	45.6
Wilting point (% v)	9.4	8.3	7.4	7.3

3.2 Stage-specific Variable Decomposition and Feature Construction

To facilitate interventional analysis, raw agroenvironmental records were transformed into a set of causal features. The primary objective was to operationalize water availability as a series of growth-stage-specific treatments, enabling the model to isolate temporally distinct sensitivities.

Additionally, lagged yield and pre-planting climate features were constructed to explicitly control for historical productivity and antecedent confounding conditions.

3.2.1 Stage-specific Decomposition: Start-, Mid-, and Late-season

Evaluating the timing of water sensitivity requires partitioning the single seasonal aggregate into agronomically meaningful treatment windows. The early-season (May–June 30) encompasses early crop establishment and initial canopy development. The mid-season (July 1–August 14) corresponds to the critical tuber initiation and active bulking phase. The late-season (August 15–September 30) captures tuber maturation and senescence prior to harvest. Let $s \in \{\text{ES}, \text{MS}, \text{LS}\}$ denote these early-, mid-, and late-season windows, respectively. For every postal code i and year t, climate variables were independently aggregated to define the stage-specific water availability treatments $T_{i,t}^{\text{ES}}$, $T_{i,t}^{\text{MS}}$, and $T_{i,t}^{\text{LS}}$.

3.2.2 Crop Water Deficit/Excess Estimation

Within each temporal window, water availability was quantified as a continuous deficit/excess variable, calculated as the net difference between accumulated precipitation and the crop's physiological evapotranspiration demand. The foundational atmospheric water demand was quantified using the standardized FAO-56 Penman–Monteith equation for reference evapotranspiration (ET_0) [49], which robustly integrates thermal, radiation, and aerodynamic drivers:

$$\text{ET}_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{3.1}$$

where R_n is net radiation at the crop surface, G is soil heat flux density, T is mean daily air temperature at 2 m height, u_2 is wind speed at 2 m height, $(e_s - e_a)$ represents the vapor pressure deficit, Δ is the slope of the vapor pressure curve, and γ is the psychrometric constant.

Specific crop evapotranspiration (ET_c) was then derived by scaling this reference demand dynamically with a stage-specific crop coefficient K_c:

$$ET_c = ET_0 \times K_c \tag{3.2}$$

To accurately align this calculation with site-specific phenology, planting dates were estimated for each postal-code–year unit within a May 1–31 window using rigorous temperature thresholds: (i) a mean daily temperature $\geq 7°C$ for two consecutive days, and (ii) an absence of projected frost (minimum temperature $<$ -2°C) over the subsequent seven days. These estimated planting dates initialized the accumulation of growing degree days (GDD_7) using a 7°C base temperature, reflecting the biological baseline for potato development [50]:

$$GDD_7 = \max\left(T_{\text{mean}} - 7°C,\ 0\right) \tag{3.3}$$

Rather than applying fixed calendar-based stage coefficients, K_c was dynamically parameterized by accumulated GDD_7 to reflect the crop's true phenological progress:

- Initial ($GDD_7 < 300$): $K_c = 0.70$
- Development ($300 \leq GDD_7 < 600$): $K_c = 0.90$
- Mid-Season ($600 \leq GDD_7 < 900$): $K_c = 1.15$
- Late-Season ($GDD_7 \geq 900$): $K_c = 0.85$

The final water deficit/excess treatment for each postal code i, year t, and stage s was computed as:

$$T^s_{i,t} = P^s_{i,t} - ET_{c}{}^s_{i,t} \tag{3.4}$$

where $P^s_{i,t}$ is accumulated precipitation. Negative values indicate a crop water deficit, while positive values represent a water surplus.

Descriptive statistics of these constructed features (Fig. 4) confirm a stark seasonal contrast across PEI. Early- and late-season periods generally exhibited water surpluses (postal-code means of 25.1 mm and 25.9 mm, respectively). Conversely, the mid-season was characterized by persistent, severe water deficits, averaging −55.0 mm globally. Region CW experienced the sharpest mid-season stress (mean: −72.1 mm), followed by Regions CE (−53.9 mm), NW (−37.9 mm), and east (−25.5 mm). This distribution reinforces the agronomic premise that the mid-season tuber bulking phase is the primary period of exposure to moisture stress.

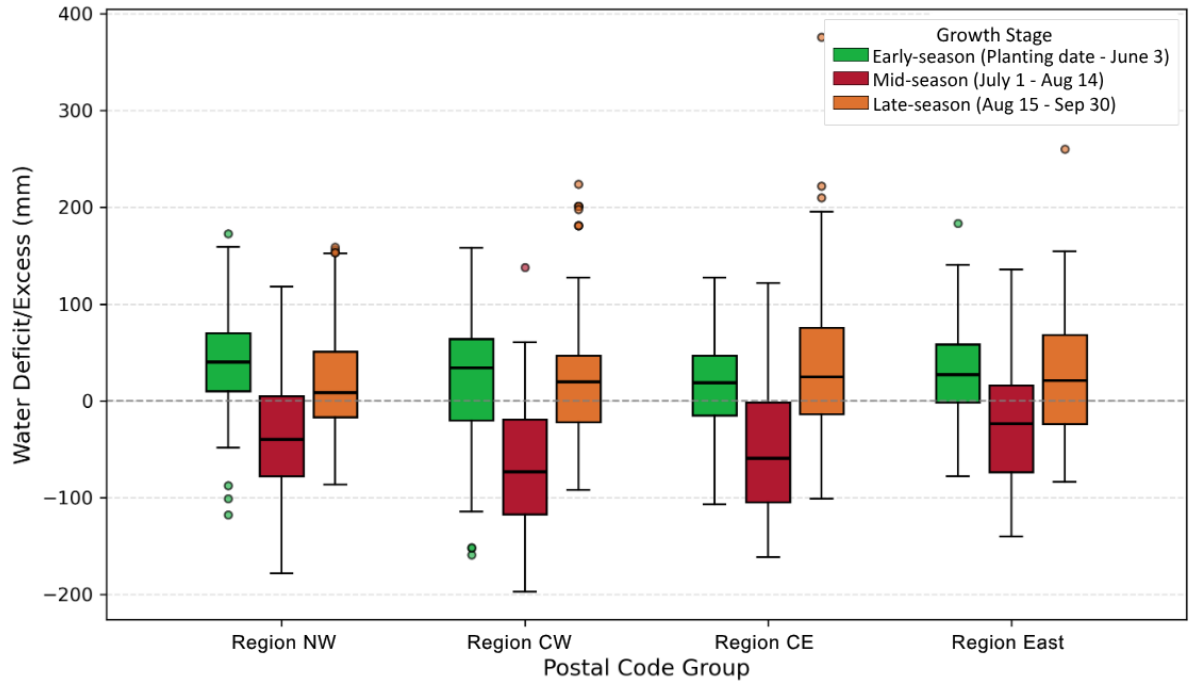

Figure 4: ***Stage-specific water deficit/excess by postal-code grouped region in PEI.*** *Boxplots illustrating inter-annual dataset variability (1982–2024) of stage-specific water deficit/excess, highlighting consistent mid-season deficits. Boxes denote the interquartile range (IQR), internal lines represent the median, and whiskers extend to 1.5×IQR.*

3.2.3 Yield Persistence Encoding and Pre-planting Climate Controls

Crop yield commonly exhibits strong temporal persistence because agricultural productivity is shaped by environmental, biological, technological, and management factors that carry over across years [51]. At the postal-code level, these persistent factors may include local soil

quality, cultivar and management practices, pest and disease pressure, and long-term productivity trends. If such historical yield dependence is ignored, model residuals may retain spatial and temporal autocorrelation, reducing predictive reliability and potentially overstating the precision of estimated water–yield relationships.

To account for this persistence without introducing temporal information leakage, historical productivity was encoded using only yield information available prior to the target year. Two lagged yield features were constructed for each postal-code–year observation. The first was the immediate prior-year yield:

$$Y_{i,t}^{\text{lag1}} = Y_{i,t-1},$$

where $Y_{i,t-1}$ denotes the potato yield in postal code i during the previous year. The second was a lagged three-year moving average:

$$Y_{i,t}^{\text{MA3}} = \frac{1}{3}\left(Y_{i,t-1} + Y_{i,t-2} + Y_{i,t-3}\right). \tag{3.5}$$

These variables summarize short-term and medium-term yield persistence while ensuring that the model is trained only on information that would have been available before the observed yield outcome in year t. In this way, lagged yield features help represent the baseline productivity capacity of each postal code and reduce residual variation associated with persistent local conditions that are not fully captured by measured soil and climate variables.

Pre-planting climate controls were also included to represent antecedent environmental conditions before the stage-specific water availability treatments occurred. In particular, accumulated precipitation before planting was treated as a pre-treatment covariate because it may influence both subsequent water deficit/excess and final yield. Together, lagged yield features, pre-planting climate variables, and soil properties provide a set of adjustment and prognostic covariates that improve model stability while preserving the temporal ordering required for causal dose-response

estimation. The complete set of variables utilized in the causal framework is detailed in Table 5 (Appendix).

3.3 Causal Design and Adjustment Set

To estimate the causal effect of water availability on potato yield, we define water deficit/excess as the treatment variable (T) and potato yield as the outcome (Y). Because the data are observational, causal interpretation requires explicit identification of the confounding structure linking water availability, soil conditions, historical productivity, and yield.

We adopt the Structural Causal Model (SCM) paradigm [19, 52], where causal assumptions are represented as a Directed Acyclic Graph (DAG). In this DAG, nodes represent variables, and directed edges ($X \rightarrow Y$) assert that X is a direct cause of Y, with no directed cycles permitted. The causal estimand of interest is the effect of T on Y, which we estimate by constructing a valid adjustment set Z through constrained causal discovery to learn the graph structure, followed by formal backdoor-criterion-based selection.

The study employed causal discovery to learn the causal structure directly from data using conditional independence (CI) tests [52]. Specifically, we applied the PC algorithm [53], augmented with domain-informed prior rules to prevent agronomically implausible edge orientations.

The PC algorithm proceeds in two phases:

(1) Skeleton Discovery: The algorithm begins with a fully connected undirected graph. It tests conditional independence $A \perp\!\!\!\perp B \mid C$ for adjacent pairs, iterating over conditioning sets C of increasing size. If independence holds, the edge is removed, and C is recorded as the separating set.

(2) Orientation: The skeleton is then partially oriented using three rules [52]: (1) identifying v-structures (unshielded colliders) $A \rightarrow B \leftarrow C$; (2) preventing new colliders by orienting

remaining edges away from existing ones; and (3) preventing directed cycles.

If residual cycles remain, one edge per cycle is pruned to yield a valid partially directed acyclic graph. To ensure the learned graph respects physical and temporal domain knowledge, we imposed the following agronomic constraints prior to orientation:

(i) Outcome Constraint: Only yield (Y) may be caused by water deficit/excess (T).

(ii) Sink Constraint: No variable can be caused by the current year's yield (Y).

(iii) Historical Constraint: Historical yield features cannot cause contemporary, non-outcome environmental variables.

(iv) Temporal Constraint: Directed edges among seasonal variables must strictly respect chronological order (e.g., Early Season conditions may cause Late Season conditions, but not vice versa).

(v) Exogenous Constraint: No localized agronomic covariate is permitted to cause broad, exogenous meteorological or soil drivers.

These priors significantly reduce the CI search space and guarantee a domain-appropriate structure before data-driven orientation.

From the resulting DAG, valid adjustment sets were identified using the backdoor criterion [19]. An adjustment set Z satisfies this criterion when conditioning on Z blocks every non-causal path entering the treatment T through an incoming arrow, without conditioning on descendants of T or opening paths through colliders [52]. DoWhy [54] was used to enumerate valid adjustment sets and to distinguish treatment–outcome confounders from prognostic covariates that influence only the outcome.

The final stage-specific DAG is presented in Figure 5. Pre-planting precipitation is a confounder because it directly influences both stage-specific water deficit/excess and potato yield. For the mid- and late-season analyses, prior-stage water deficit/excess is an additional confounder because antecedent water conditions may influence both the water balance observed during

the current growth stage and the final yield outcome. Consequently, the minimal adjustment set for the early-season treatment contains pre-planting precipitation, whereas the minimal adjustment sets for the mid- and late-season treatments contain both pre-planting precipitation and the corresponding prior-stage water deficit/excess. Conditioning on these variables blocks the identified backdoor paths between the stage-specific treatment and potato yield.

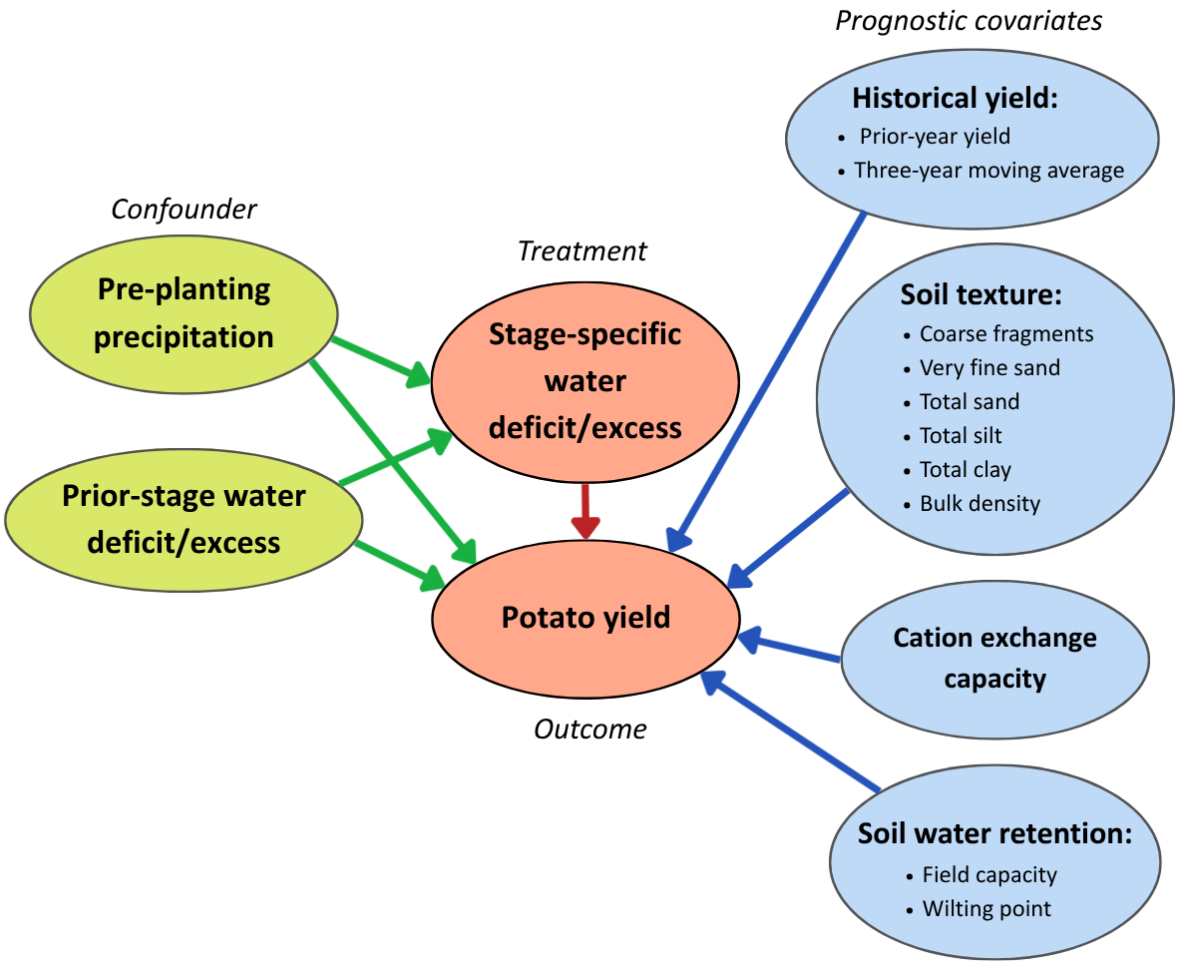

Figure 5: ***Final directed acyclic graph used to identify the causal effect of stage-specific water deficit/excess on potato yield.*** *The orange nodes represent the treatment and outcome. Green nodes and arrows represent treatment–outcome confounders: pre-planting precipitation and, where applicable, prior-stage water deficit/excess, each of which influences both the current-stage treatment and potato yield. The red arrow represents the causal pathway from stage-specific water deficit/excess to yield. Blue nodes and arrows represent pre-treatment prognostic covariates that directly predict yield but do not cause the current-stage treatment, including historical yield, soil texture, cation-exchange capacity, and soil water-retention characteristics. Prior-stage water deficit/excess is not applicable to the first analyzed growth stage.*

In addition to the confounders required for causal identification, the S-learner included a broader

set of pre-treatment prognostic covariates. These variables comprised prior-year yield, the three-year moving average of yield, soil-texture properties, bulk density, cation-exchange capacity, field capacity, and wilting point. In the final DAG, these variables have directed paths to potato yield but no directed paths to the current-stage treatment. They are therefore not required to block the identified backdoor paths. Nevertheless, including them in the outcome model can explain additional variation in yield, improve predictive precision, and stabilize estimation of the conditional dose–response function, provided that they are measured before the treatment and are not descendants of the treatment [19].

A critical assumption underlying this entire strategy is no unmeasured confounding, which dictates that all variables influencing both treatment assignment and the outcome are fully captured in the dataset. Without randomized experiments, this assumption is theoretically required to isolate true causal effects from mere correlation [55].

3.3.1 Positivity Diagnostics and Common Support

Estimating a causal dose–response function for a continuous treatment requires sufficient empirical support across the treatment range of interest. In this study, the positivity assumption requires that, for each observed covariate profile Z_i, the conditional treatment density $f_{T|Z}(t \mid z_i)$ is non-negligible at the treatment levels where causal effects are estimated [26, 37, 40]. When particular water deficit/excess levels are rarely observed for specific soil–climate–yield histories, the estimator must rely on extrapolation rather than comparable empirical observations, weakening the causal interpretation of the estimated dose–response curve.

Although the main analysis estimates stage-specific water–yield responses, positivity was diagnosed and handled using total growing-season water availability. Specifically, the aggregate

treatment was defined as the sum of early-, mid-, and late-season water deficit/excess:

$$T^{GS} = T^{ES} + T^{MS} + T^{LS}. \tag{3.6}$$

This aggregate growing-season water deficit/excess variable provides a practical common-support screen for the overall water regime in which stage-specific counterfactual curves are later estimated. Consequently, the resulting causal estimand should be interpreted within the retained growing-season overlap population.

To diagnose and address potential positivity violations, we estimated the generalized propensity score (GPS), defined as the conditional density of the observed treatment given covariates, $\hat{f}_{T|Z}(T_i \mid Z_i)$. Because T^{GS} is continuous, the GPS was estimated using an out-of-sample residual kernel density estimation approach. First, the expected treatment $\hat{T}(Z)$ was modeled with a CatBoost regressor. To preserve the empirical treatment distribution and avoid underrepresenting rare treatment regions, the treatment model was trained using five-fold StratifiedKFold cross-validation based on treatment bins, with bin-balanced sample weights. This produced out-of-sample predictions $\hat{T}_i$ for each observation.

The out-of-sample residuals were then computed as

$$e_i = T_i - \hat{T}_i, \tag{3.7}$$

and a one-dimensional kernel density estimator was fitted to the residual distribution, denoted by $\hat{f}_E(e)$. Under the additive-noise approximation $T \approx \hat{T}(Z) + E$, the point GPS for each observation was estimated as

$$GPS_i = \hat{f}_{T|Z}(T_i \mid Z_i) \approx \hat{f}_E(e_i). \tag{3.8}$$

In addition to point GPS values, the treatment range was divided into contiguous intervals $I_j = (a_j, b_j]$ to evaluate counterfactual support across the treatment axis. For each observation i,

the conditional probability of falling within interval I_j was approximated as

$$p_{ij} = \Pr(T \in I_j \mid Z_i) \approx \int_{a_j}^{b_j} \hat{f}_E\left(t - \hat{T}_i\right)\, dt. \tag{3.9}$$

The integral was approximated numerically using the trapezoidal rule by evaluating the KDE density at equally spaced points within each interval.

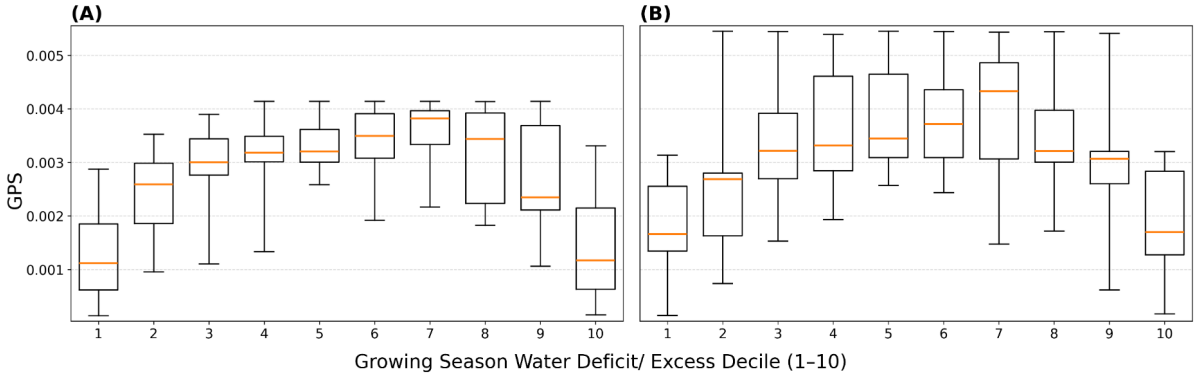

Figure 6: ***GPS distributions by growing-season water deficit/excess decile before and after trimming.*** *Panel (A) shows the generalized propensity score (GPS) distributions for the full sample before trimming, grouped by realized growing-season water deficit/excess decile. Panel (B) shows the corresponding GPS distributions after GPS-based trimming and refitting the treatment model on the retained sample. Boxes represent the interquartile range, orange lines indicate medians, and whiskers show the non-outlying range. Lower GPS values indicate weaker conditional support for the observed treatment level given the covariate profile.*

Common support was evaluated using two complementary diagnostics. First, GPS boxplots by treatment decile were used as a realized-support check. This diagnostic summarizes how likely each unit's observed treatment value is, conditional on its covariates. In Figure 6A, the full-sample GPS distributions show substantial heterogeneity across treatment deciles, with lower GPS values appearing near the boundaries of the treatment range. This indicates that some observations, particularly in extreme deficit or excess regimes, have weak empirical support even though they are present in the observed data. However, because deciles aggregate relatively wide portions of the continuous treatment range, this diagnostic alone cannot precisely identify where support collapses.

Second, interval-based range analysis was used as a counterfactual support map. Rather than evaluating only each unit's realized treatment value, this diagnostic examines the estimated

interval probabilities p_{ij} across the treatment grid. Intervals with systematically low probability mass identify regions where counterfactual predictions would rely on weak empirical support. To mitigate positivity violations, observations with inadequate support were removed using a two-part trimming rule based on the integrated interval probabilities. For each observation i, the own-bin probability was defined as the probability assigned to the interval containing the observed treatment value:

$$p_i^{\text{own}} = p_{i,j(i)}, \tag{3.10}$$

where $j(i)$ denotes the interval containing T_i. To ensure that retained observations also had local neighborhood support, a second criterion was defined using the adjacent ± 1 interval window:

$$p_i^{\text{local}} = \sum_{j \in \{j(i)-1,\, j(i),\, j(i)+1\}} p_{ij}. \tag{3.11}$$

An observation was retained only if both support criteria exceeded data-adaptive thresholds anchored at the fifth percentile:

$$p_i^{\text{own}} \geq \tau_{\text{own}} \quad \text{and} \quad p_i^{\text{local}} \geq \tau_{\text{local}}. \tag{3.12}$$

This trimming procedure removed 28 of 526 observations, corresponding to 5.32% of the dataset. Removed observations were concentrated primarily in the extreme treatment deciles, with approximately 26% removed from both Decile 1 and Decile 10. This pattern confirms that overlap was weakest at the outer boundaries of the growing-season water deficit/excess distribution.

After trimming, the GPS model was refitted on the retained sample using the same out-of-sample procedure. The post-trimming diagnostics indicate improved common support. In Figure 6B, the GPS distributions are more stable across deciles, and the near-zero support observations present in the full sample are largely removed. While this does not eliminate weaker support at the treatment boundaries, it substantially reduces structural overlap violations and confines subsequent dose–response estimation to a better-supported empirical population.

Therefore, the trimming protocol changes the target estimand from the full observational population to the retained overlap population. Within this population, estimated dose–response curves are supported by more credible empirical comparisons and are less dependent on extrapolation into poorly observed water deficit or excess regimes.

3.4 S-learner Estimation and Model Validation

The causal dose–response function was estimated using an S-learner framework [16, 36]. In this approach, a single supervised regression model is trained to approximate the conditional expected yield,

$$\mathbb{E}[Y \mid T, Z],$$

where T denotes the stage-specific water deficit/excess treatment and Z denotes the adjustment covariates identified from the causal graph. The dataset was split temporally into a training period from 1982 to 2016 and a held-out test period from 2017 to 2024. Four candidate base learners were evaluated: Random Forest, XGBoost, CatBoost, and Ridge regression.

Models were trained on the GPS-trimmed training data from 1982 to 2016, so that the estimated outcome surface was learned from observations satisfying the empirical overlap condition. Specifically, trimming was applied before causal curve estimation to reduce the influence of postal-code–year observations whose observed treatment values were conditionally unsupported given their covariates. As a result, the fitted outcome surface $\hat{f}(Z,T)$ was less likely to be driven by extrapolation in weakly supported regions of the treatment space. This design makes the resulting dose–response estimates overlap-qualified and reduces the risk that estimated yield responses to seasonal water deficit/excess reflect artifacts of model extrapolation rather than empirically supported comparisons.

Cross-validation was performed using postal-code group-aware folds, which ensured that obser-

vations from the same postal code did not appear simultaneously in the training and validation folds. This strategy helps prevent spatial data leakage because repeated observations from the same postal code may share persistent management, and productivity. Hyperparameters were first tuned using RandomizedSearchCV, a randomized hyperparameter-search procedure that samples a fixed number of parameter combinations from predefined distributions rather than evaluating all possible combinations. This approach provides an efficient first-stage search across a broad hyperparameter space. The best-performing region of the search space was then refined using targeted GridSearchCV, which exhaustively evaluates all specified combinations within a narrower parameter grid.

Model performance was evaluated on the untrimmed held-out test period from 2017 to 2024 using mean absolute error (MAE) and root mean squared error (RMSE). Based on the held-out evaluation results in Table 3, the Random Forest model was selected as the final base estimator for dose–response analysis. The purpose of this untrimmed evaluation was predictive rather than causal: it assessed whether the trained model could generalize to realistic future years, including rare or extreme water conditions that may occur in practice. However, causal interpretation of the dose–response curves remains restricted to the overlap-supported treatment domain identified by the GPS diagnostics and trimming procedure. Therefore, the untrimmed held-out evaluation provides an honest test of predictive robustness under real field conditions, whereas the causal conclusions and irrigation guidance are reported only for the overlap-qualified region where the data provide sufficient empirical support.

Table 3: Held-out predictive performance of candidate S-learner base models for potato yield prediction.

Model	**MAE** (t/ac)	**RMSE** (t/ac)
Ridge regression	1.83	2.25
Random Forest	**0.97**	**1.30**
XGBoost	1.30	1.75
CatBoost	1.36	1.78

After selecting the final estimator, postal-code-level dose–response curves were generated by varying the treatment T over a grid of 42 equally spaced values spanning the trimmed treatment range. For each postal code p, the adjustment covariates $\bar{Z}_p$ were fixed at their within-postal-code mean values, which allowed the model to predict yield under hypothetical water deficit/excess levels while holding local soil, climate, and historical productivity conditions constant. The model was then evaluated under interventions on T to obtain

$$\hat{\mu}_p(t) = \hat{f}(\bar{Z}_p, t), \qquad t \in t_{\text{grid}}. \tag{3.13}$$

This standardization procedure produces a location-specific dose–response curve for each postal code and provides the basis for comparing heterogeneous water–yield responses across regions. Causal robustness was evaluated using a placebo refutation test. In this test, the treatment variable was randomly shuffled within postal codes to destroy the original treatment–outcome association while preserving the overall data structure. The S-learner was then re-estimated using the placebo treatment. If the original dose–response curves reflect a meaningful causal signal rather than a modeling artifact, the placebo curves should appear relatively flat and unstructured compared with the original estimates.

3.5 Marginal Yield Response and Effective Water-allocation Ranges

To identify the water deficit/excess range where additional water availability provides the greatest yield return, marginal yield response was extracted from each stage-specific dose-response curve. For each postal code p, the estimated dose-response curve $\hat{\mu}_p(t)$ was first smoothed using a Savitzky–Golay filter, a local polynomial smoothing method that fits a low-degree polynomial within a moving window [56]. Unlike a simple moving average, this filter reduces numerical noise while better preserving local curve features such as slopes, peaks, and turning points. The

marginal response was then approximated using a centered finite-difference derivative:

$$\left.\frac{d\hat{Y}}{dT}\right|_{t=t_k} \approx \frac{\hat{\mu}_p(t_{k+1}) - \hat{\mu}_p(t_{k-1})}{t_{k+1} - t_{k-1}}.$$

This quantity represents the expected change in potato yield, measured in $\mathrm{t/ac}$, associated with a one-millimetre increase in water availability at treatment level t_k. Positive values indicate yield gains from additional water availability, whereas values near zero indicate limited incremental benefit. Negative values indicate yield penalties associated with further increases in water availability, which may occur under excess-water conditions.

Effective water-allocation ranges were defined as treatment intervals where the marginal response exceeded 0.01 $\mathrm{t/ac/mm}$. These intervals identify the portions of the dose-response curve where additional water availability is expected to generate agronomically meaningful yield gains. The maximum marginal response and the width of the effective range were then summarized across postal codes and growth stages to compare when water interventions are most likely to improve yield.

Chapter 4: Results

The empirical results are organized around the stage-specific causal dose–response estimates and their irrigation-relevant interpretation. First, the estimated dose–response curves are compared across early-season, mid-season, and late-season water deficit/excess to assess whether potato yield responds differently depending on the timing of water availability. The corresponding marginal-return heatmaps identify the treatment ranges where additional water availability produces positive, negligible, or negative yield responses. Placebo refutation results are then used to evaluate whether the estimated curves retain meaningful treatment-related signal when compared with shuffled-treatment models. Because the mid-season treatment exhibits the strongest and most consistent response pattern, the subsequent analysis focuses on mid-season marginal yield response and effective water-allocation ranges, which provide the basis for the discussion of supplemental irrigation prioritization and soil-related response heterogeneity.

4.1 Stage-specific Dose–response Curves and Placebo Refutation

The estimated stage-specific dose–response curves reveal clear differences in how potato yield responds to water deficit/excess across the growing season. Figures 7–9 show regional aggregated dose–response curves and the corresponding marginal-return heatmaps for early-season (ES), mid-season (MS), and late-season (LS) water deficit/excess. In each figure, Panel A reports the predicted yield response across the water deficit/excess gradient, while Panel B reports the local marginal return, defined as the change in predicted yield associated with an additional millimetre of water availability.

For early-season water deficit/excess, the common-support treatment range extends approximately from deficit conditions to moderate excess-water conditions. Across regions, the estimated curves indicate that predicted yield is maximized near balanced water availability, where ES water deficit/excess is close to 0 mm (Fig. 7A). Yield generally increases as early-season deficit is relieved, but the response is moderate and region-specific. The marginal-return heatmap (Fig. 7B) shows that positive marginal returns are concentrated mainly within moderate deficit intervals rather than across the full treatment range. On average, the effective early-season water-allocation range produces a marginal return of approximately 0.016 t/ac/mm. Among the four regions, Region CW shows the steepest early-season gain, with a marginal return of approximately 0.026 t/ac/mm when ES water deficit is relieved from about -50 to -40 mm. However, the early-season curves also indicate that additional water becomes less beneficial once water balance approaches or exceeds zero. This excess-water penalty is most evident in Region NW, where the predicted curve begins to decline after moving from balanced conditions into surplus conditions, with an estimated total yield loss of approximately 1.0 t/ac when early-season water excess reaches about 100 mm. Region CW also shows a negative response under surplus conditions, with an estimated yield loss of approximately 0.5 t/ac as early-season water excess increases from about 44 to 100 mm. In contrast, Regions CE and Region east show only mild negative responses under early-season water excess. Overall, early-season water availability contributes to yield formation, but its effect is less uniform and less pronounced than the mid-season response. This pattern is consistent with the interpretation that early-season water stress may be partly compensated for if later-season water availability is favorable, while excess water at the early season can still create localized penalties through delayed emergence, poor drainage, or restricted early root development.

The mid-season curves show the strongest and most consistent water–yield response among the three growth stages (Fig. 8A). The common-support range for MS water deficit/excess

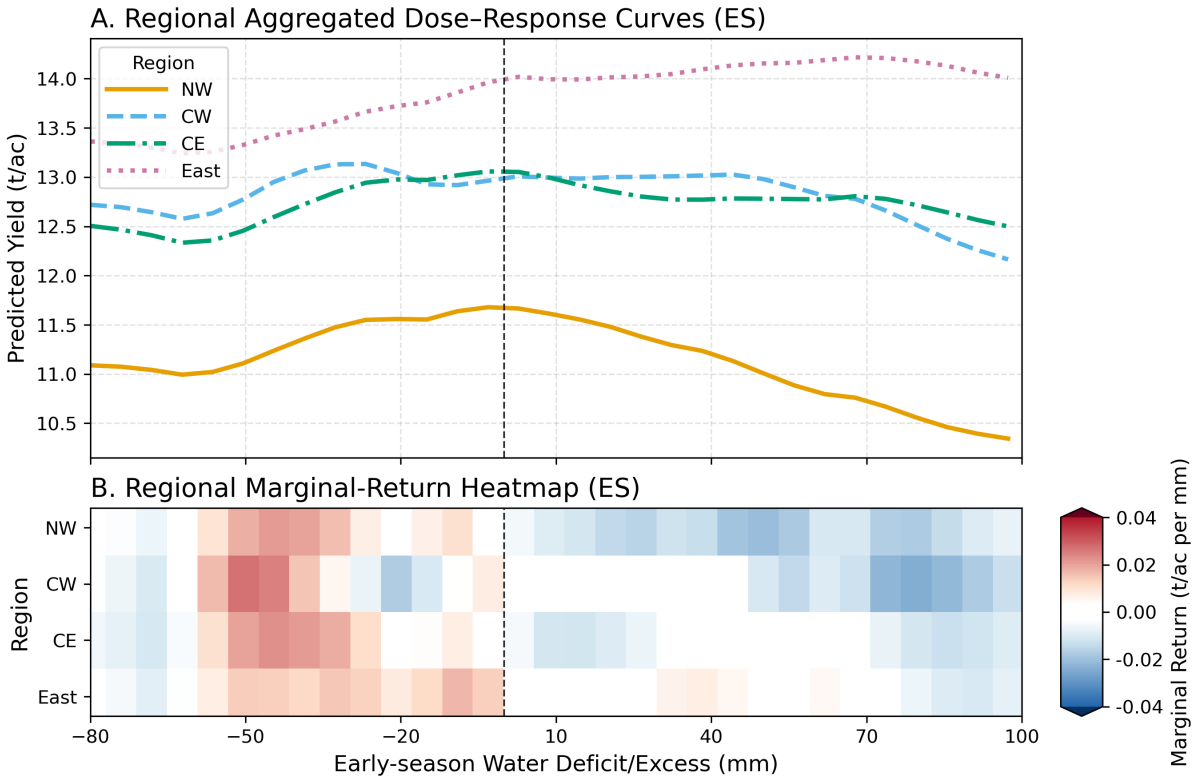

Figure 7: ***Regional early-season dose–response curves and marginal returns.*** *Panel A shows regional aggregated dose–response curves for early-season water deficit/excess. The vertical dashed line denotes balanced water availability, where deficit/excess equals zero. Panel B shows the corresponding marginal return across the treatment grid; red indicates positive yield return from additional water availability, blue indicates negative return, and near-white values indicate limited marginal response.*

extends from approximately −170 to 80 mm. Across regions, predicted yield increases sharply as mid-season water deficit is reduced from severe deficit conditions toward approximately −100 to −70 mm. Beyond this range, the curves generally flatten, indicating diminishing marginal returns once the most severe mid-season water stress is relieved. This response pattern differs from the early-season curves because the yield gain is not only visible in the dose–response curves but also consistently reinforced by the marginal-return heatmap.

The marginal-return heatmap (Fig. 8B) confirms that the strongest positive marginal returns are concentrated within the severe-to-moderate deficit interval, especially between approximately −150 and −100 mm. This interval represents the portion of the treatment space where additional water availability produces the largest expected yield gains. Regional summaries further show

that the effective range, defined as the interval where marginal return exceeds 0.01 t/ac/mm, differs across regions. Region NW shows an average marginal return of 0.026 t/ac/mm over −137 to −100 mm. Region CW shows a lower average marginal return of 0.017 t/ac/mm, but over a wider interval from −167 to −100 mm. Region CE exhibits a similar average marginal return to Region NW, 0.027 t/ac/mm, over −152 to −100 mm. Region east has the highest average marginal return, 0.046 t/ac/mm, over a narrower interval from −152 to −108 mm, corresponding to an expected total yield gain of approximately 2.7 t/ac within that effective range.

These differences indicate that mid-season deficit relief is beneficial across all regions, but the magnitude and width of the effective response range are spatially heterogeneous. Region east appears most responsive to additional water within a relatively narrow deficit interval, whereas Region CW exhibits a broader but weaker response. Therefore, both the marginal return and the width of the effective water-deficit interval should be considered when translating the dose–response curves into irrigation-allocation priorities. A region with a high marginal return may provide a strong yield response to targeted water application, while a region with a wider effective range may benefit across a larger set of deficit conditions.

Late-season water deficit/excess displays a different response pattern from the mid-season curves. The common-support range for LS water deficit/excess extends from approximately −60 to 150 mm. Across regions, the maximum predicted yield generally occurs when LS water deficit/excess is near 0 mm, indicating that balanced late-season water availability is associated with the most favorable yield outcome (Fig. 9A). However, compared with mid-season, the yield gain from deficit relief is weaker. Region NW shows the clearest positive late-season response, with an expected total yield gain of approximately 0.9 t/ac as LS water deficit is relieved toward balanced conditions. The other regions show only mild total yield gains, generally less than 0.5 t/ac, when late-season deficit is relieved.

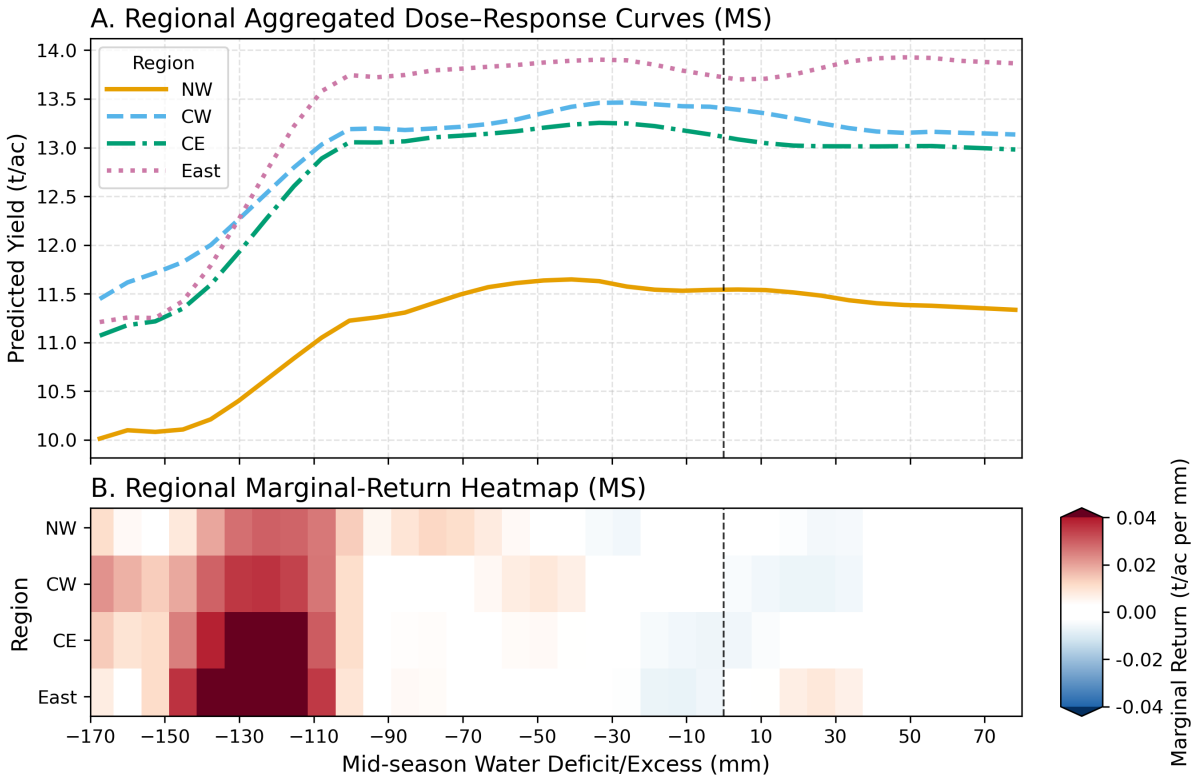

Figure 8: ***Regional mid-season dose–response curves and marginal returns.*** *Panel A shows regional aggregated dose–response curves for mid-season water deficit/excess. Panel B shows local marginal returns across the same treatment range. Strong positive marginal returns are concentrated under severe-to-moderate deficit conditions, indicating the water-balance interval where additional water availability provides the greatest expected yield gain.*

The marginal-return heatmap (Fig. 9B) shows weaker positive returns under deficit conditions and more pronounced negative returns under surplus conditions. In particular, negative marginal returns become evident when late-season water excess approaches high surplus levels, especially around 100 mm and above. This pattern suggests that late-season water availability is more closely associated with penalties from excessive moisture than with strong yield gains from deficit relief. In practical terms, late-season water management appears to be less about maximizing additional water input and more about maintaining balanced conditions while avoiding water excess. This is consistent with the agronomic expectation that excessive late-season moisture may increase risks related to poor drainage, disease pressure, or bad harvest condition.

Taken together, the three stage-specific dose–response results show that water availability affects potato yield differently depending on growth stage. Table 4 summarizes the main response

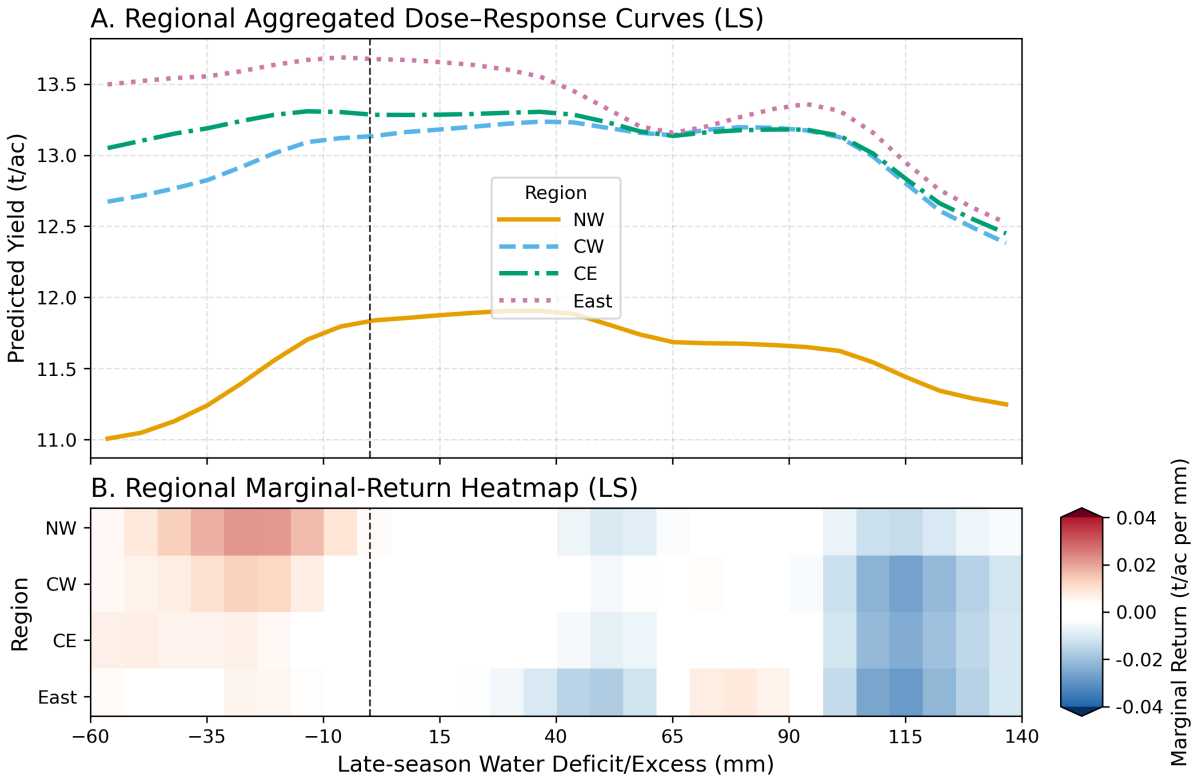

Figure 9: ***Regional late-season dose–response curves and marginal returns.*** *Panel A shows regional aggregated dose–response curves for late-season water deficit/excess. Panel B shows the corresponding marginal-return heatmap. Positive marginal returns are weaker than in the mid-season case, while negative returns become more evident under late-season excess-water conditions.*

characteristics across ES, MS, and LS water deficit/excess, including the effective response range, relative response strength, and irrigation-relevant interpretation.

Table 4: Summary of stage-specific water-yield response characteristics.

Stage	Main effective range	Response strength	Main interpretation
Early-season	Moderate deficit near balance	Mild to moderate	Maintain balance; avoid excess
Mid-season	Approximately −150 to −100 mm	Strongest	Priority range for water allocation
Late-season	Deficit relief toward balance	Weak	Avoid excess-water penalty

The placebo refutation test provides additional evidence that the estimated water–yield relationships capture stage-specific treatment signals rather than only model-induced artifacts. Figure 10 compares the response span of the real dose–response curves with curves estimated after shuffling the treatment within postal codes. Response span is defined as the vertical range of the estimated dose–response curve and therefore summarizes the magnitude of yield variation implied by the

treatment response.

Across all three stages, the observed stage-specific water deficit/excess treatment curves exhibit larger response spans than their placebo counterparts, indicating that the original treatment variable contains predictive structure not retained after randomization. The separation is largest for mid-season water deficit/excess, where the median real response span is 1.68 t/ac, compared with 0.48 t/ac for the placebo curves. This corresponds to a reduction of approximately 71% after treatment shuffling. Early-season and late-season curves also show real–placebo separation, but the differences are smaller. Together, the dose–response curves, marginal-return heatmaps, and placebo refutation results indicate that mid-season water status contains the clearest yield-forming signal among the three growth stages.

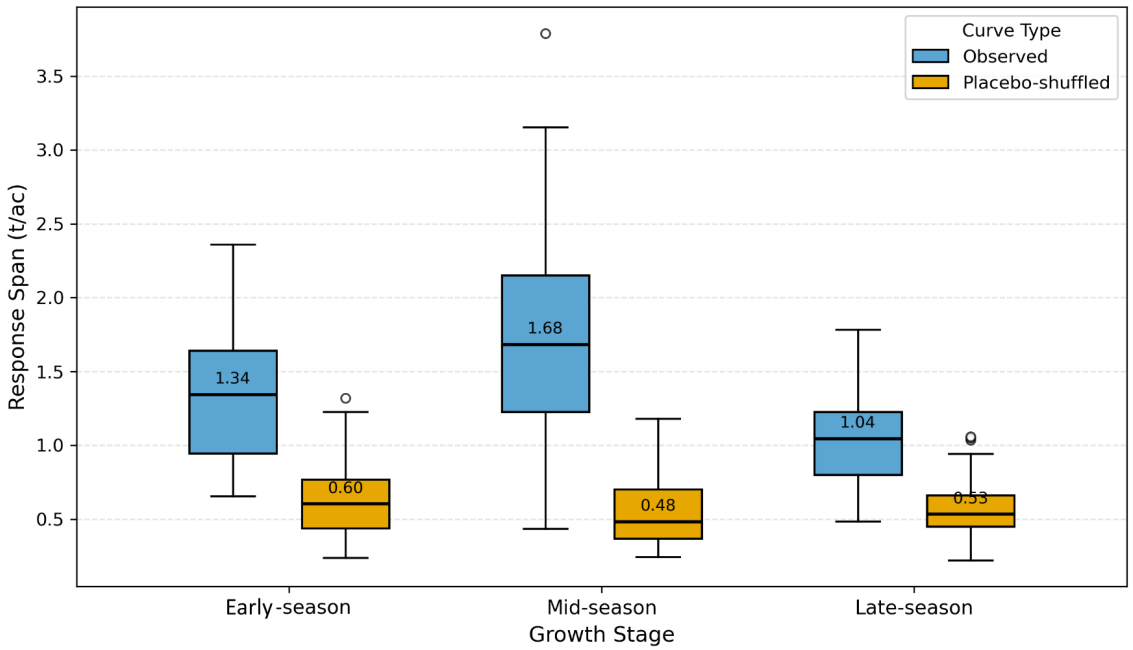

Figure 10: ***Placebo refutation based on observed and placebo-shuffled-treatment response spans.*** *Response span is defined as the vertical range of an estimated dose-response curve. Blue boxes summarize postal-code response spans from the real treatment models, while yellow boxes summarize placebo treatment spans generated by shuffling water deficit/excess values within postal codes, thereby preserving postal-code structure while disrupting the original treatment–yield alignment. Numbers inside boxes denote median response span. Larger real-versus-placebo separation indicates stronger evidence that the estimated curve reflects treatment related signal rather than random model structure.*

4.2 Marginal Yield Response and Effective Water-allocation Ranges

The marginal-return analysis provides a more decision-oriented interpretation of the estimated stage-specific dose–response curves. While the dose–response curves show how predicted yield changes across the full water deficit/excess gradient, the marginal-return curves identify where an additional millimetre of water availability is expected to generate the greatest yield benefit. This distinction is important because the results do not suggest that additional water is equally valuable across all water conditions or all growth stages. Instead, the value of additional water depends strongly on both timing and baseline water stress.

Across the early-season and late-season results, the effective water-management strategy is primarily to maintain water availability near balance and avoid excess moisture. In the early-season, positive marginal returns occur under moderate deficit conditions, but the total response range is relatively small and surplus water can generate yield penalties, particularly in Region NW. In the late-season, deficit relief produces only limited yield gains, while negative marginal returns become more pronounced under excess-water conditions across regions. These patterns suggest that ES and LS water conditions are relevant to yield, but their estimated response magnitudes are generally modest, often below approximately 1 t/ac. Therefore, early-season and late-season water interventions may be most useful as risk-prevention measures, especially for avoiding water excess, rather than as the primary source of yield improvement.

In contrast, the mid-season marginal-return results identify a clearer and more agronomically meaningful water-allocation target. The strongest marginal returns occur when mid-season water deficit is relieved within the severe-to-moderate deficit range, approximately from -150 to -100 mm. Within this interval, additional water availability is associated with the largest expected yield gains. Across regions, mid-season deficit relief within the effective water-allocation range produces an average yield gain of approximately 2 t/ac, with the strongest regional response

observed in Region east. Once water availability approaches less severe deficit conditions, the marginal return declines, indicating that further water input becomes less productive after the main stress threshold has been relieved.

This result provides an irrigation-relevant interpretation of the causal dose–response estimates. Rather than prioritizing water application uniformly across the growing season, the findings suggest that supplemental irrigation or water-allocation planning should focus first on fields or regions experiencing substantial mid-season deficit. The effective water-allocation range is therefore not the entire treatment range, but the portion of the mid-season deficit region where marginal response remains positive and agronomically meaningful. From a decision-support perspective, the estimated curves indicate that water allocation would be most efficient when used to relieve severe mid-season deficit toward approximately -100 mm, while avoiding unnecessary water application once marginal returns begin to flatten.

Chapter 5: Discussion

The results demonstrate that potato yield response to water availability is strongly stage-dependent. Among the three growth-stage windows considered in this study, mid-season water deficit/excess exhibits the largest and most consistent dose–response signal, the strongest concentration of positive marginal returns, and the clearest separation between real and placebo response spans. This finding indicates that the timing of water availability is central to yield formation: the same amount of water deficit or excess does not have the same agronomic meaning when it occurs at different points in the growing season.

The pronounced mid-season response is agronomically plausible, as this period overlaps with tuber initiation, bulking, and peak yield accumulation. During this stage, water stress directly limits canopy function, photosynthetic activity, and the movement of assimilates into developing tubers [57]. Consequently, relieving a mid-season deficit produces a stronger yield response than similar relief during less sensitive periods. In contrast, early-season water availability shows moderate but less consistent yield sensitivity. This reflects the importance of early moisture for crop establishment while recognizing that early-season stress may be partially compensated for if later conditions are favorable [58]. Late-season water availability shows weaker evidence of yield gain from additional water and a sharper penalty under excess moisture, indicating that late-season management is primarily about avoiding moisture-related losses rather than boosting yield potential.

Overall, the results support the central argument of this thesis: causal dose–response estimation provides a more decision-relevant understanding of agricultural water response than either predic-

tive modeling or drought monitoring alone. Predictive models can identify variables associated with yield, and drought indicators can identify areas experiencing water stress, but neither directly indicates where additional water is expected to produce the greatest marginal yield gain. By estimating stage-specific dose–response curves and marginal returns, the proposed framework connects water stress exposure with expected yield response, allowing water-allocation decisions to be interpreted in terms of both risk and return.

5.1 Implications for Supplemental Irrigation Allocation and Optimization

The mid-season marginal-response results can be translated into a postal-code-level irrigation decision-support framework by jointly considering the expected yield benefit of deficit relief and the frequency with which severe water deficits occur. Figure 11 illustrates the spatial heterogeneity of these two complementary dimensions across PEI. Panel A distinguishes postal codes where additional water within the effective deficit-relief range is associated with larger marginal yield gains, whereas panel B identifies postal codes that have historically experienced severe mid-season water deficits more frequently. Considered together, the maps provide spatial context for supplemental irrigation allocation because areas with high estimated marginal returns may offer greater yield benefits when irrigation is applied, while areas with frequent deficit exposure may face a more persistent need for intervention. However, neither indicator alone is sufficient to determine irrigation relevance: postal codes may exhibit high responsiveness but infrequent exposure, or frequent exposure but comparatively modest yield returns.

To integrate these two dimensions directly, Figure 12 compares postal codes according to the observed frequency of mid-season water deficit below -100 mm and the average marginal return within the effective deficit-relief range from -150 to -100 mm. This joint representation distinguishes locations that frequently experience severe mid-season water stress from those

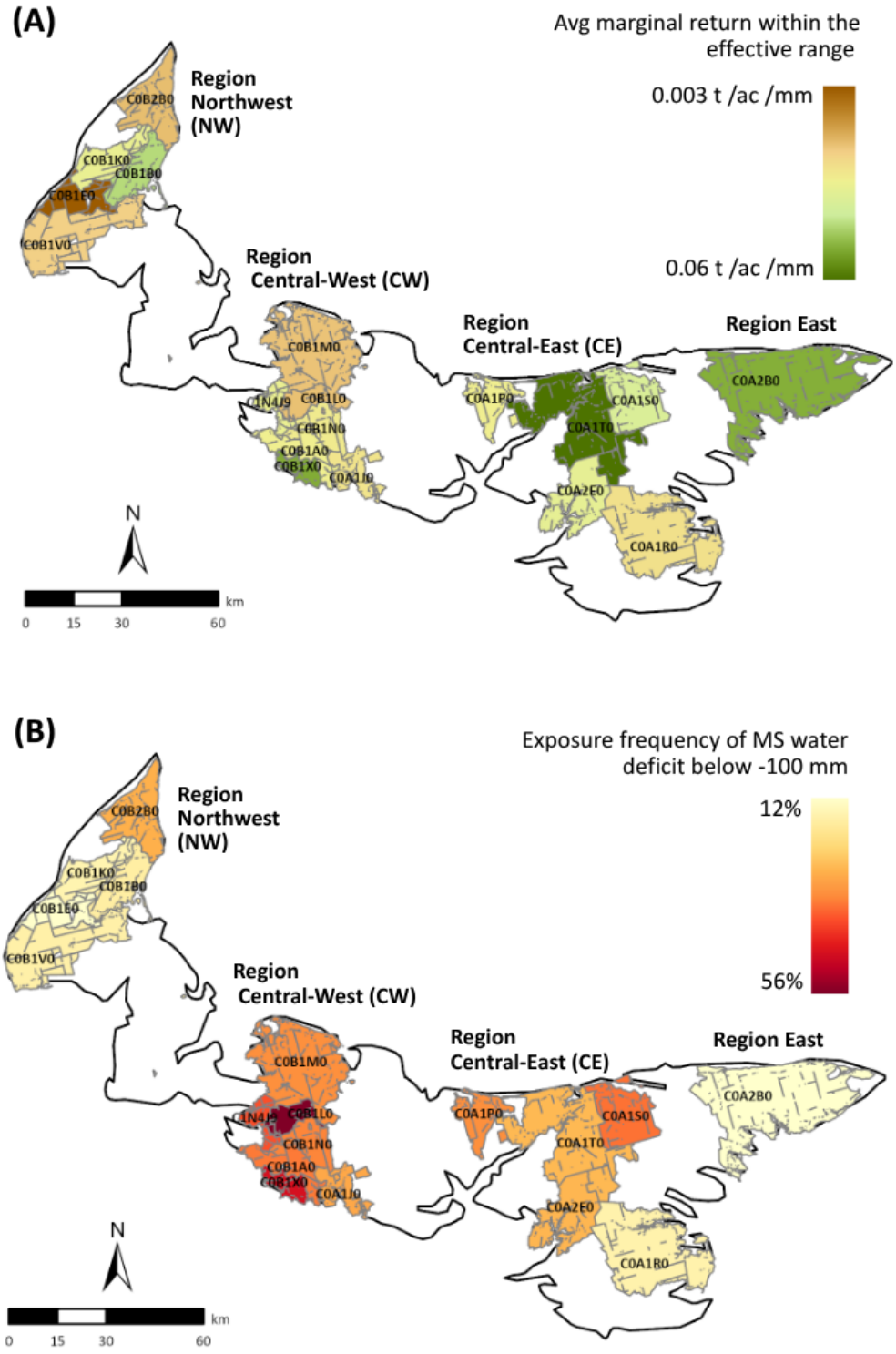

Figure 11: ***Postal-code-level spatial distribution of mid-season marginal yield return and severe water-deficit exposure across PEI.*** *Panel A shows the average marginal yield return within the effective mid-season deficit-relief range from* -150 *to* -100 *mm, as derived from the estimated postal-code-level dose–response curves; greener shading indicates a larger estimated yield return per additional millimetre of water. Panel B shows the historical exposure frequency of severe mid-season water deficit, calculated as the percentage of postal-code–year observations with water deficit/excess below* -100 *mm; darker red shading indicates more frequent exposure.*

in which relieving that stress is expected to generate comparatively large marginal yield gains. Consequently, the combined risk–return perspective provides a more informative basis for identifying where supplemental irrigation may be most relevant, while also supporting comparisons among locations with different combinations of deficit exposure and yield responsiveness.

For interpretation, the decision-reference thresholds are fixed at 25% observed exposure frequency and 0.025 t/ac/mm average marginal return. Postal codes above both thresholds are interpreted as having higher irrigation opportunity because they combine meaningful exposure to mid-season deficit with above-threshold expected yield response. These thresholds are not intended as fixed agronomic rules, but as a decision-support screening device for distinguishing different irrigation-planning contexts, where supplemental irrigation may be more likely to provide agronomically meaningful returns.

The postal-code-level plot shows that irrigation opportunity is not evenly distributed across PEI. High irrigation opportunity is concentrated mainly in the CE and CW regions. Four of the five postal codes in CE, or approximately 80%, fall in the high irrigation opportunity quadrant. Similarly, about 70% of postal codes in CW fall at or above the decision-reference thresholds. These postal codes combine relatively frequent exposure to mid-season deficit with above-threshold marginal return, suggesting that they may offer strong potential for supplemental irrigation because they face meaningful deficit risk and are expected to respond positively when water stress is relieved.

Several postal codes show frequent mid-season deficit exposure but more modest marginal response. Postal codes C0B 1L0 and C0B 1M0 in CW, together with C0B 2B0 in NW, fall below the marginal-return threshold while remaining above the exposure-frequency threshold. For these locations, irrigation may still be relevant because severe mid-season deficit occurs relatively frequently. However, the expected yield gain per additional millimetre of water is lower than in the high-opportunity quadrant. As a result, irrigation investment in these postal codes should be

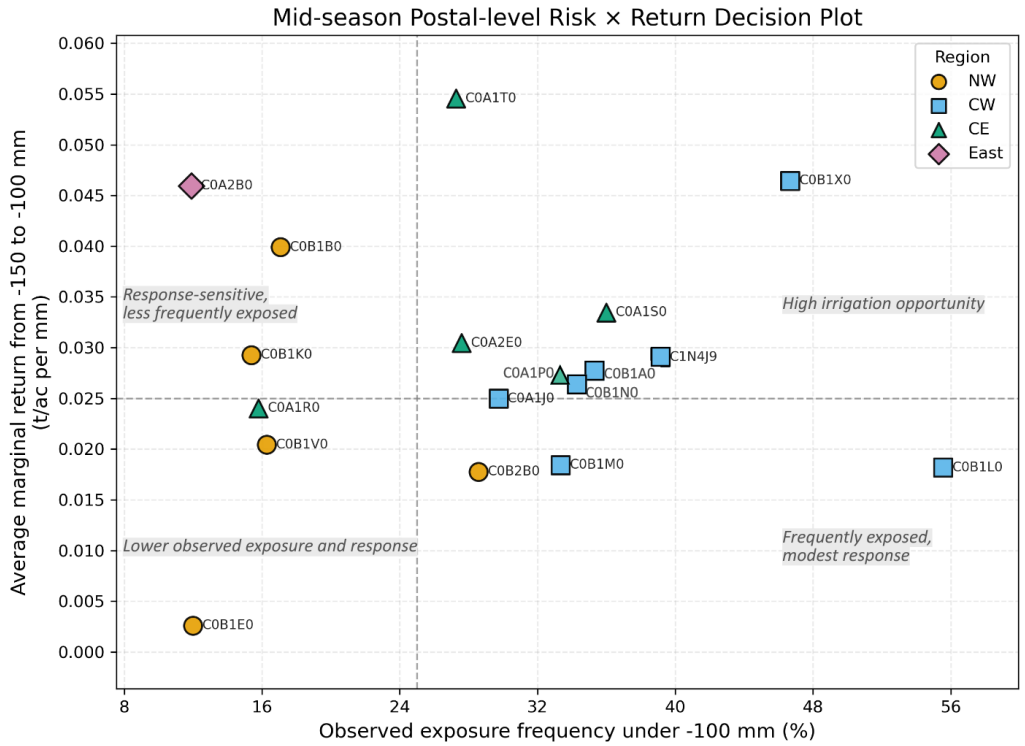

Figure 12: ***Postal-code-level mid-season risk–return irrigation decision plot.*** *Each point represents one postal code, with colors indicating regional grouping. The x-axis shows the observed frequency of mid-season water deficit below* -100 *mm, representing historical exposure to severe mid-season deficit. The y-axis shows the average marginal return within the effective deficit-relief range from* -150 *to* -100 *mm, representing the expected yield response to additional water availability. Dashed reference lines denote the decision thresholds used in this study:* 25% *exposure frequency and* 0.025 t/ac/mm *average marginal return.*

evaluated carefully against water availability, infrastructure requirements, pumping cost, and expected economic return.

A different pattern is observed for postal codes with strong marginal response but less frequent severe-deficit exposure. Postal code C0A 2B0 in Region East shows a pronounced average marginal return but relatively low observed frequency of mid-season deficit below -100 mm. In addition, approximately 40% of postal codes in Region NW fall into this response-sensitive but less frequently deficit-exposed quadrant. These postal codes may be highly responsive to supplemental water during drought years, but they may not require irrigation as consistently as postal codes that experience severe mid-season deficit more frequently. For these locations, event-triggered or drought-year irrigation planning may be more appropriate than broad irrigation

expansion.

Finally, postal codes C0B 1V0 and C0B 1E0 in Region NW, together with C0A 1R0 in Region CE, fall below both decision-reference thresholds. These postal codes have both lower observed deficit exposure and lower expected marginal return within the effective deficit-relief range. As a result, they represent comparatively lower irrigation opportunity for broad supplemental irrigation expansion under the historical conditions analyzed in this study. This does not imply that irrigation would never be useful in these locations, but rather that they show weaker evidence of frequent deficit exposure and strong marginal yield response compared with other locations. The postal-code-level risk–return plot supports a targeted rather than uniform irrigation strategy. The highest expected benefits arise not simply where water deficits occur, but where deficit exposure and marginal yield response coincide. The postal-code-level analysis provides a finer screening layer by identifying within-region variation in irrigation opportunity. This is important because postal codes within the same broader region can differ in both historical deficit exposure and estimated yield responsiveness. The framework therefore provides information that is complementary to drought monitoring alone: it identifies not only where water stress is likely, but also where relieving that stress is expected to generate the greatest marginal yield response. These findings advocate for a targeted irrigation strategy where resource allocation targets the intersection of deficit exposure and maximum marginal yield response. Because water, infrastructure, and labour are finite, effective decision-support systems must prioritize the expected yield gain from reducing a deficit, not just its presence. The highest expected benefits arise not simply where water deficits occur, but where deficit exposure and marginal yield response coincide. In this sense, the causal dose-response framework provides information that is complementary to drought monitoring alone: it identifies not only where water stress is likely, but also where relieving that stress is expected to generate the greatest yield response. Such information can support future irrigation optimization models by defining priority regions,

effective treatment ranges, and stage-specific response functions.

5.2 Model Explainability and the Distinction between Prediction and Intervention

The Shapley Additive exPlanations (SHAP) summary analysis provides insight into how the S-learner leverages covariates to predict yield (Fig. 13). Crucially, SHAP values reflect predictive contribution—conditional on the fitted model and data structure [59]—rather than causal effects. The causal interpretation relies entirely on the prior identification strategy, positivity handling, and placebo refutation. SHAP serves strictly as an explainability tool to confirm that the predictive model heavily weights the variables central to our causal analysis.

As shown in Fig. 13, MS Water Def Excess is the most influential feature, aligning perfectly with the dose–response results showing mid-season water availability as the strongest driver of yield. Its wide SHAP distribution, capturing both positive and negative contributions, reflects the non-linear dose–response curves: water's impact depends entirely on whether the crop is experiencing severe deficit, optimal balance, or excess moisture.

Historical yield variables, including prior-year yield (Lag Yield) and three-year moving average (Moving Avg), also rank highly in the SHAP summary. These features are important for prediction because they capture temporal persistence in postal-code yield patterns. Prior-year yield and lagged rolling averages likely summarize persistent differences in soil productivity, management regimes, local growing conditions, and other slow-moving factors that are not fully observed in the dataset. However, these variables should be interpreted as baseline predictors rather than intervention levers. A producer cannot directly intervene on last year's yield to increase current yield. Instead, lagged yield features help the model account for persistent productivity differences so that the estimated water-response relationship is not forced to absorb

all historical variation across postal codes.

This highlights a critical distinction for agricultural decision support: predictive models prioritize variables with the largest predictive weight, but causal frameworks require actionable variables. Mid-season water availability is uniquely valuable because it is both highly predictive and actionable via supplemental irrigation or drainage. Ultimately, the SHAP analysis confirms that the model grounds its predictions in the primary causal treatment rather than relying solely on historical persistence.

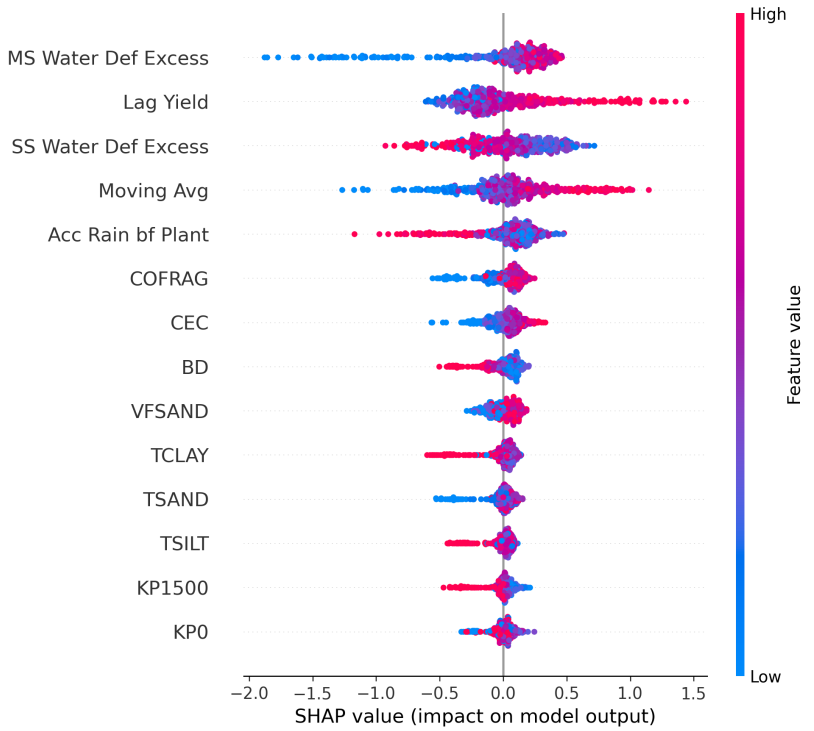

Figure 13: ***SHAP summary for the mid-season S-learner fitted on the trimmed training set.*** *Each point represents one observation. The x-axis shows the SHAP value, which indicates the contribution of a feature to the model prediction relative to the model baseline. Positive SHAP values increase predicted yield, while negative SHAP values decrease predicted yield. Features are ranked on the y-axis by mean absolute SHAP value. Point color represents the feature value, with blue indicating low values and red indicating high values. The figure should be interpreted as a model-explainability diagnostic rather than as causal evidence; causal interpretation comes from the identified dose–response framework.*

The SHAP results also contextualize the adjustment variables. Soil properties, accumulated rainfall before planting, early-season water deficit/excess, and yield-persistence variables all contribute to the model prediction. Their inclusion helps adjust for background confounding

that may influence both water availability and yield. While predictive, they are not all direct intervention targets; soil texture represents a slow-changing constraint, while pre-planting rainfall sets the seasonal context. The practical value of SHAP here is explaining how the S-learner synthesizes this multifaceted environmental context to estimate yield response.

5.3 Soil-related Heterogeneity in Mid-season Response Curves

The spatial heterogeneity observed in the mid-season response curves suggests that local soil conditions influence how yield responds to water deficit relief. Figure 14 summarizes the Spearman correlations between soil properties and three curve-derived heterogeneity indicators: effective range width, average marginal return within the effective range, and maximum predicted yield. These indicators describe different aspects of the estimated response curves. Effective range width measures how broad the water-deficit interval is where marginal return remains above the selected threshold. Average marginal return measures the intensity of yield response within that effective range. Maximum predicted yield summarizes the highest yield level predicted along the mid-season water-balance curve.

These correlations should be interpreted as exploratory associations rather than causal soil effects. Soil properties were not treated as experimental treatments in this analysis, and the correlations are based on curve-derived summaries. Therefore, the heatmap should not be read as evidence that changing one soil property would necessarily cause a specific change in water response. Instead, it provides evidence about which soil characteristics are associated with stronger or weaker estimated water-response patterns across postal-code regions, helping explain why a uniform water deficit produces varying yield responses across regions.

Texture and drainage-related variables show the clearest associations with mid-season response heterogeneity. Total sand is positively associated with average marginal return within the

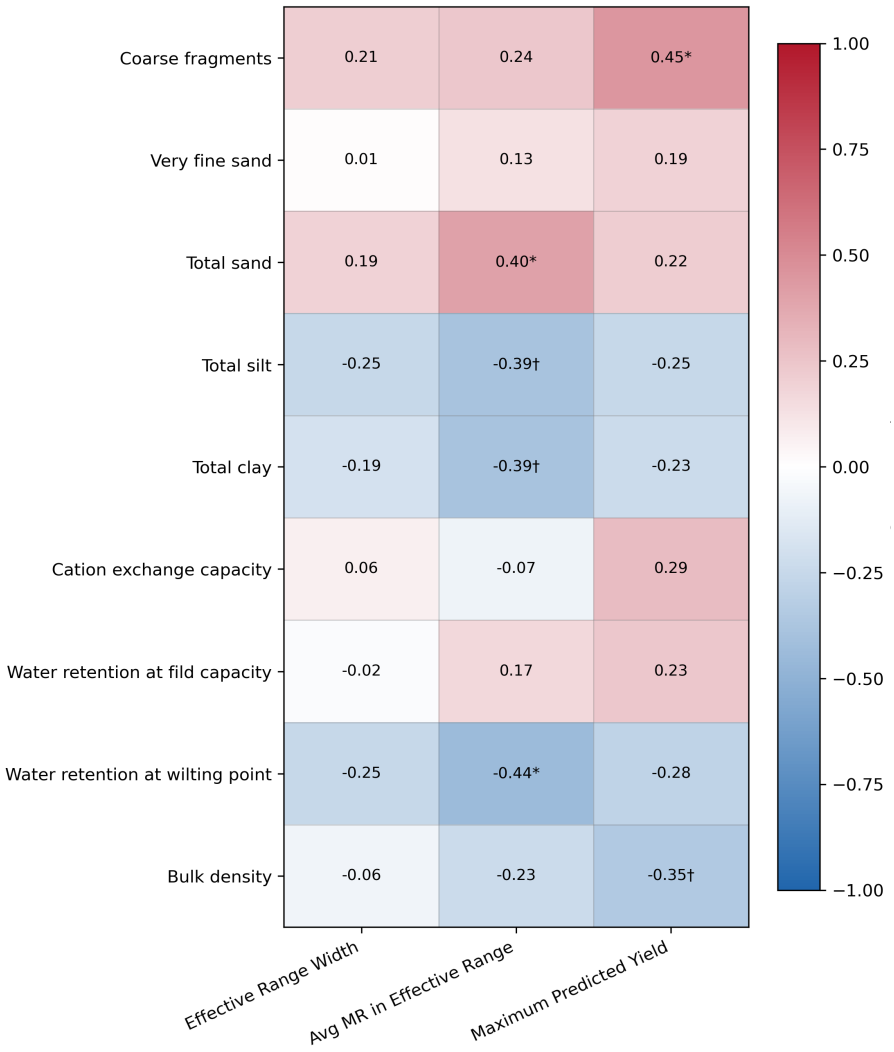

Figure 14: ***Associations between soil properties and mid-season response-curve indicators.*** *Cells report Spearman correlation coefficients between soil variables and indicators extracted from postal-code mid-season dose–response curves. The three curve-derived indicators are effective range width, average marginal return within the effective range, and maximum predicted yield. Red indicates positive association and blue indicates negative association. Significance symbols denote the reported correlation test levels:* ** $p < 0.01$, * $p < 0.05$, *and* † $p < 0.10$. *These associations are exploratory and should not be interpreted as causal soil effects.*

effective range, with a Spearman correlation of approximately 0.40. This suggests that sandier soils may show stronger yield response when mid-season deficit is relieved. One plausible explanation is that lighter-textured or better-drained soils have lower water-holding capacity and therefore experience stronger water limitation during dry mid-season periods. In such settings, supplemental water may more directly relieve stress and translate into yield gains.

In contrast, total silt and total clay are negatively associated with average marginal return, both with correlations around -0.39. Water retention at wilting point also shows a negative association with average marginal return, approximately -0.44. Finer-textured soils with higher water retention buffer against mid-season deficits, shrinking the marginal benefit of added water. However, this buffering capacity can become a liability, increasing vulnerability to poor drainage or limited root-zone aeration during wet periods.

The heatmap also suggests that soil properties are associated with maximum predicted yield. Coarse fragments show a positive correlation with maximum predicted yield, approximately 0.45, while bulk density shows a negative association, approximately -0.35. The negative association with bulk density is consistent with the interpretation that compacted soils may restrict root growth, water movement, and aeration, thereby limiting attainable yield. The positive association with coarse fragments should be interpreted more cautiously because coarse fragments can reflect multiple underlying soil conditions, including drainage, rooting environment, and landscape position. Rather than implying that coarse fragments directly increase yield, this association suggests that the soil environments represented by higher coarse-fragment values may coincide with higher predicted yield potential in the observed data.

These soil-heterogeneity results reinforce that irrigation allocation must be soil-informed. Water deficit relief appears most valuable where soil conditions allow additional water to be converted into yield gains without creating drainage or root-zone limitations. Conversely, soils with higher fine-texture content or greater water retention at wilting point may buffer some deficit conditions

but may also reduce the marginal benefit of additional water.

5.4 Application Guidelines for Decision Support

This proposed framework functions as a decision-support layer, not a rigid irrigation prescription. The effective deficit-relief range identifies a target zone for planning, but real-world execution must still account for forecasted rainfall, supply constraints, and economic viability.

In practice, recommended irrigation amounts must be offset by anticipated rainfall to avoid pushing fields into excess-moisture penalties. If a field is in a severe mid-season deficit, the gross water requirement is the difference between the current deficit and the target effective range boundary. Forecasted rainfall must be deducted from this gross requirement. Because the dose–response curves clearly show diminishing returns and excess-water penalties, over-application is both economically inefficient and agronomically risky.

The framework also requires that the adjustment variables used for prediction reflect the actual conditions of the decision period. In the experimental analysis, covariates were fixed at local historical averages to construct interpretable postal-level response curves. This is appropriate for estimating and comparing standardized response patterns. However, in operational use, the covariate profile should represent the real-time field, season, or forecast scenario. For example, current accumulated rainfall before planting, current or forecasted stage-specific water deficit/excess, and recent yield history should be used where available. Using outdated or overly averaged covariate values could produce recommendations that do not reflect the actual risk conditions of the year being managed.

A practical implementation of the framework could follow four steps: (1) forecast the current mid-season water deficit for the target field; (2) identify if the field is expected to enter the high-return deficit range, especially the severe-to-moderate deficit interval around -150 to -100 mm; (3)

combine this information with regional risk–return priority and soil context to determine whether irrigation is likely to be beneficial; (4) calculate irrigation need after subtracting forecasted rainfall and considering local constraints such as available water, system capacity, soil drainage, and cost. This workflow would allow the causal dose–response estimates to inform irrigation planning while still respecting the operational realities of farm-level decision-making.

Ultimately, management interpretations vary by stage. For early- and late-season windows, the priority is maintaining balance and avoiding excess moisture. For the mid-season, the objective is prioritizing deficit relief while marginal returns remain high. This stage-specific interpretation replaces the oversimplified conclusion that more water is always better with a targeted strategy identifying precisely when and where water delivers the greatest agronomic value.

5.5 Limitations and Future Directions

This thesis establishes an end-to-end causal machine-learning framework for estimating stage-specific water–yield dose–response relationships from long-term observational agricultural data. By integrating causal identification, positivity diagnostics, flexible outcome modeling, marginal-response estimation, and spatial decision-support analysis, the study provides a foundation from which several new research problems can now be addressed. The following limitations define the appropriate scope of the present findings while also identifying opportunities made possible by the framework developed in this thesis.

The causal interpretation of the estimated dose–response curves depends on the assumption of no unmeasured confounding. Conditional on the identified adjustment set, water deficit/excess is assumed to be independent of potential yield outcomes. The adjustment set incorporates pre-treatment climate conditions, soil characteristics, and yield-persistence variables; however, unobserved factors may still introduce residual bias if they influence both water availability

and yield. These factors may include cultivar selection, fertilizer application, planting density, pest and disease pressure, supplemental irrigation history, and other within-field management practices that vary across postal codes or years. The estimated dose–response curves should therefore be interpreted as causal under the stated identification assumptions rather than as effects obtained from a randomized experiment.

Nevertheless, the causal identification procedure developed in this thesis provides a structured basis for progressively incorporating these additional management variables. Because the framework explicitly separates confounders, outcome predictors, treatments, and outcomes, future studies can introduce newly available farm-management information without redesigning the overall analytical workflow. This creates an opportunity to evaluate how management practices modify water–yield relationships and to determine whether the stage-specific responses identified here remain stable after controlling for more detailed production decisions.

A second limitation concerns spatial aggregation. The analysis is conducted at the postal-code level, with yield, climate, and soil information harmonized to a common spatial unit. This aggregation enabled the construction of a consistent multi-decadal dataset and supported the estimation of geographically differentiated response curves across Prince Edward Island. However, it also smooths over field-level differences in slope, drainage, soil texture, management intensity, and access to irrigation. The resulting curves are therefore most appropriate for regional planning, resource-prioritization, and decision-support screening rather than direct field-level irrigation prescriptions.

Because this thesis demonstrates that causal water–yield responses and marginal irrigation returns can be differentiated spatially even within a relatively small agricultural region, the framework can now be extended to field- and farm-level applications. Incorporating georeferenced yield records, higher-resolution soil measurements, remotely sensed crop indicators, and farm-management data would allow the regional patterns identified here to be tested at finer

spatial scales. Such an extension could reveal within-postal-code heterogeneity, support field-specific response estimation, and determine how accurately regional priority patterns translate into individual irrigation decisions.

The GPS-based trimming results identify another important boundary of the analysis. Positivity-driven exclusion of weakly supported observations showed that historical records contain insufficient examples of certain extreme deficit and surplus conditions for some covariate profiles. Trimming strengthens causal credibility by restricting estimation to treatment levels supported by comparable historical observations, but it also limits inference for water-balance conditions that rarely occurred in the observed record. This issue is particularly relevant under climate change, as future growing seasons may experience combinations of water deficit, excess moisture, and temperature stress that fall outside the historical common-support region.

At the same time, explicitly identifying the empirical support boundary is itself an important contribution of this thesis. Rather than allowing the estimator to produce unconstrained predictions across all conceivable water conditions, the proposed workflow distinguishes evidence-supported causal estimation from extrapolation. This diagnostic capability provides a principled starting point for studying conditions beyond the historical record.

One promising extension is Knowledge-Guided Machine Learning (KGML) [60,61]. Agronomic and biophysical knowledge could be incorporated into model architectures, loss functions, parameter constraints, or post-estimation rules. For example, future models could be constrained to avoid biologically implausible yield increases under water deficits approaching the permanent wilting point or to respect attainable yield limits imposed by root-zone soil conditions. These constraints would not replace the need for empirical support or justify causal claims outside the observed treatment range. However, because the present framework identifies where empirical support ends, KGML could be applied selectively to produce more physically and agronomically defensible scenario projections under unprecedented climate conditions.

The marginal-response functions developed in this thesis also create opportunities to move beyond response estimation toward irrigation optimization. The present analysis identifies where additional water is expected to generate the greatest marginal yield benefit and where frequent deficit exposure coincides with strong yield responsiveness. Because these marginal returns have now been quantified across growth stages and postal codes, they can serve as agronomic inputs to economic and multi-objective optimization models.

Future work could combine the estimated response functions with pumping costs, water availability, irrigation-system capacity, energy prices, crop prices, and risk preferences. This would allow the framework to progress from identifying biologically responsive conditions to recommending economically viable water-allocation strategies. Multi-objective optimization could further examine trade-offs among expected yield, irrigation cost, water-use efficiency, production risk, and environmental impact. In this way, the causal estimates produced by the thesis provide the response surface needed to investigate operational irrigation decisions rather than merely documenting historical associations.

Finally, the finding that water–yield relationships differ substantially across crop-development stages opens several biological and climatic research directions. The present study identifies mid-season water availability as the most influential stage-specific treatment for the study population, while also showing that early- and late-season responses follow different patterns. Building on this result, future studies can evaluate whether the timing and magnitude of these responses differ among potato cultivars, maturity classes, soil environments, or production systems.

The framework can also be expanded from a single continuous water-balance treatment to compound climate exposures. Interactions among water deficit, excess moisture, heat stress, solar radiation, soil drainage, and antecedent growing conditions may produce responses that cannot be represented by separate one-dimensional curves. Because the current thesis establishes the identification, support-diagnostic, and estimation components required for continuous-treatment

analysis, future research can extend these components to multivariate or sequential treatments. Such analyses could investigate how stress accumulation and timing jointly shape yield and could provide more comprehensive decision support for climate-adaptive potato production.

In summary, the contribution of this thesis extends beyond identifying a single water threshold or irrigation range. It establishes a causal, stage-specific, spatially differentiated, and support-aware framework for translating observational agricultural records into decision-relevant water–yield response information. The results identify mid-season deficit relief as the strongest irrigation opportunity within the supported historical range, reveal substantial geographic variation in both deficit exposure and marginal return, and show that soil conditions help explain heterogeneity in estimated responses. Model-interpretability analysis further clarifies how environmental and historical-yield variables contribute to the fitted predictions, although it does not independently establish causality. Together, these achievements provide the methodological and empirical foundation for higher-resolution estimation, management-aware causal analysis, economically optimized irrigation allocation, cultivar-specific evaluation, and climate-scenario modeling beyond historically observed conditions.

Chapter 6: Conclusion

This thesis began from a practical problem in agricultural decision support: crop yield models are increasingly accurate, but accuracy alone does not answer the management questions that producers, planners, and water-resource decision makers face. A predictive model can estimate whether yield is likely to be high or low, but it does not necessarily explain whether changing a controllable condition would improve the outcome. In water-sensitive crop systems, this distinction is central. The relevant question is not only whether water availability is associated with yield, but how yield would respond under alternative levels of water deficit or excess, when that response is strongest, and where water intervention would be most valuable.

The central research question of this thesis asked how a causal machine learning dose–response framework can estimate spatially and seasonally heterogeneous water–yield relationships in agriculture and translate these relationships into credible decision support for irrigation scheduling and broader water-resource planning. The answer developed across the thesis is that such a framework is possible when predictive modeling is embedded within a broader causal workflow. This workflow must identify the relevant adjustment set, diagnose whether treatment levels are empirically supported, estimate continuous dose–response curves, evaluate the robustness of those curves, and translate the resulting marginal responses into management-relevant terms. In other words, causal machine learning becomes useful for agricultural decision support not simply because it fits flexible models, but because it links flexible prediction to explicit assumptions about intervention, support, and interpretation.

The main methodological contribution of this thesis is the construction of an end-to-end

continuous-treatment causal workflow for agricultural water-response analysis. The framework combines domain-informed causal identification, generalized propensity score diagnostics, common-support trimming, S-learner dose–response estimation, group-aware model evaluation, marginal-return analysis, and placebo refutation. Each component addresses a specific weakness that would arise if a standard predictive model were used alone. The causal graph and adjustment set address confounding; the generalized propensity score diagnostics address overlap and extrapolation; the S-learner captures nonlinear response patterns; the marginal-return analysis translates response curves into intervention-relevant quantities; and the placebo refutation evaluates whether the estimated curves retain treatment-related signal rather than reflecting only model-induced structure. Together, these components move the analysis from yield prediction toward credible counterfactual reasoning.

Substantively, the thesis shows that potato yield response to water availability in Prince Edward Island is not homogeneous across the growing season. The strongest and most consistent response occurs during the mid-season period. Across the estimated dose–response curves, mid-season deficit relief produces the clearest yield-forming signal, the strongest concentration of positive marginal returns, and the largest separation between real and placebo response spans. This finding supports the agronomic expectation that the main tuber formation and bulking period is especially sensitive to water stress. The results therefore indicate that water-management decisions should not treat all growing-season water deficit as equivalent. Timing matters: a millimetre of deficit relief during the most sensitive growth period can carry a different yield implication than the same amount of deficit relief earlier or later in the season.

The early-season and late-season results provide an important contrast. Early-season water availability shows moderate yield sensitivity, particularly as deficit conditions move toward balance, but the response is less consistent across regions and weaker than the mid-season response. This suggests that early-season water conditions contribute to crop establishment

and yield potential, while still allowing some possibility for compensation later in the growing season. Late-season water availability shows weaker evidence of yield gains from additional water and stronger evidence of penalties under excess-water conditions. Therefore, the main late-season management implication is not necessarily to add water, but to avoid excess moisture and maintain conditions close to balance. This stage-specific pattern is one of the central empirical conclusions of the thesis: effective water management depends on matching the intervention to the crop stage, not simply maximizing water availability across the full season.

The marginal-return analysis further clarifies the decision-support value of the dose–response framework. The results do not suggest that additional water is equally beneficial across all water-balance conditions. Instead, the strongest expected gains occur when severe mid-season deficit is relieved toward a more moderate deficit range. Once water stress has been substantially reduced, marginal returns diminish. This means that the effective water-allocation range is narrower than the full observed treatment range. From a management perspective, this is a critical distinction. A dose–response curve describes the overall relationship between water availability and yield, but the marginal-return curve identifies the treatment region where intervention is most valuable. This provides a more practical basis for irrigation planning than either yield prediction or drought monitoring alone.

The thesis also shows that water-response relationships are spatially heterogeneous. Regional differences in effective ranges, average marginal returns, and maximum predicted yields indicate that the same water deficit does not imply the same yield response everywhere. Some regions experience frequent mid-season deficit but lower marginal return, while others show high marginal return but less frequent exposure. This distinction supports a risk–return interpretation of irrigation priority. Regions should not be prioritized only because they experience water stress often, nor only because they respond strongly in rare deficit conditions. The highest-priority cases are those where meaningful deficit exposure and strong expected yield response coincide. This

provides a bridge between causal estimation and resource allocation: the framework identifies not only where water stress occurs, but where relieving that stress is expected to matter most.

Soil-related heterogeneity adds another layer to this interpretation. The exploratory soil analysis suggests that texture, water retention, drainage-related properties, and root-zone physical conditions are associated with differences in water-response curves. Lighter or better-drained soil conditions appear more closely associated with stronger marginal response to mid-season deficit relief, while finer-textured or more water-retentive soils may buffer some deficit conditions but reduce the marginal benefit of additional water. These patterns should not be interpreted as causal soil effects, because soil properties were not treated as interventions in the main framework. Nevertheless, they help explain why spatial heterogeneity arises and why water-allocation decisions should be soil-informed. A field or region's response to irrigation depends not only on weather-driven water deficit, but also on whether local soil conditions allow additional water to be converted into yield gain without creating drainage or aeration limitations.

The SHAP analysis contributes a complementary insight by clarifying the relationship between prediction and intervention. Features such as historical yield variables are important for prediction because they capture persistent productivity differences across locations and years. However, they are not direct management levers. A producer cannot intervene on last year's yield to change the current season. In contrast, mid-season water availability is both predictive and actionable. Its strong contribution to the fitted model is consistent with the causal dose–response results, but the interpretation of water as an intervention comes from the causal design rather than from SHAP alone. This distinction is important for responsible use of explainable machine learning in agriculture. Model explanations can identify influential predictors, but causal reasoning is needed to determine which variables can support credible intervention decisions.

The broader contribution of this thesis is therefore twofold. First, it contributes methodologically by adapting continuous-treatment causal inference tools to an agricultural yield-modeling

problem where treatments are not naturally binary. Many agronomic inputs, including water, fertilizer, temperature exposure, pest pressure, and soil moisture, vary continuously. Treating such factors as binary indicators can simplify estimation but lose much of the structure that matters for management. By estimating dose–response curves and marginal returns, this thesis demonstrates how causal machine learning can preserve the continuous nature of agricultural interventions while still addressing confounding and support. Second, it contributes substantively by showing that PEI potato yield response to water availability is stage-specific, nonlinear, and spatially heterogeneous. These findings provide a stronger empirical foundation for targeted supplemental irrigation than a uniform water-management rule.

The decision-support implication is not that the estimated curves should be used as direct irrigation prescriptions. Rather, they should be interpreted as a causal decision layer that can inform future operational systems. In practical use, water recommendations would need to account for forecasted rainfall, current soil moisture, irrigation capacity, water availability, pumping cost, crop price, and field-level management conditions. The estimated dose–response and marginal-return curves identify where intervention is likely to be valuable, but final irrigation scheduling would require integration with real-time weather forecasts and economic constraints. Therefore, the framework developed in this thesis is best understood as a foundation for irrigation prioritization and optimization rather than a complete irrigation scheduling system.

Several limitations define the boundary of the conclusions. The analysis relies on observational data, so the causal interpretation depends on the assumption that the identified adjustment set sufficiently accounts for confounding. Unmeasured factors such as cultivar choice, fertilizer practices, disease pressure, drainage infrastructure, and field-level management may still influence both water exposure and yield. The postal-code scale also limits spatial precision. Aggregating yield, soil, and climate information to postal-code units allows long-term regional analysis, but it inevitably smooths over field-level variability. In addition, common-support trimming improves

credibility within supported treatment ranges, but it limits what can be concluded about extreme deficit or excess conditions that are rare in the historical data. These limitations do not invalidate the framework, but they clarify where the estimates are strongest and where future evidence is needed.

Future research should extend the framework in several directions. The first priority is field-level validation. Integrating farm- or field-scale yield data, soil measurements, irrigation records, cultivar information, and management practices would improve spatial resolution and reduce aggregation bias. The second priority is dynamic decision support. Future models should incorporate seasonal forecasts and within-season updates so that estimated water-response functions can support real-time irrigation decisions rather than retrospective regional analysis alone. The third priority is economic optimization. Marginal yield response should be linked with water cost, pumping cost, infrastructure constraints, and crop price so that irrigation recommendations can be evaluated in terms of expected net return rather than yield gain alone.

A further direction is to extend the causal framework to additional treatments and compound risks. Water deficit and excess are central to potato production, but they interact with temperature stress, soil constraints, disease pressure, and cultivar-specific physiology. Future work could estimate joint or multi-treatment response surfaces that better reflect the combined stresses expected under climate change. Knowledge-guided machine learning also offers a promising way to improve extrapolation beyond historical support by embedding agronomic and biophysical constraints into model structure or estimation. This would be especially useful for evaluating climate futures that may contain more frequent extreme conditions than those observed in the historical record.

In conclusion, this thesis demonstrates that causal machine learning can provide a more actionable form of agricultural yield modeling when it is designed around intervention rather than prediction alone. By estimating continuous dose–response curves for water deficit and excess, identifying stage-specific marginal returns, evaluating robustness through placebo refutation, and interpreting

spatial heterogeneity through soil and risk–return analysis, the study shows how observational agricultural data can be transformed into decision-relevant evidence. The central takeaway is that supplemental water management should be targeted, stage-specific, and context-aware. For PEI potato production, mid-season deficit relief emerges as the clearest opportunity for yield improvement, but its value depends on regional exposure, marginal response, and soil context. More broadly, the thesis offers a framework for using causal machine learning as a bridge between agricultural data science and practical climate-adaptive management.

References

[1] T. Adekanmbi, X. Wang, S. Basheer, R. A. Nawaz, T. Pang, Y. Hu, and S. Liu, "Assessing future climate change impacts on potato yields—a case study for prince edward island, canada," *Foods*, vol. 12, no. 6, p. 1176, 2023.

[2] J. Maqsood, A. A. Farooque, X. Wang, F. Abbas, B. Acharya, and H. Afzaal, "Contribution of climate extremes to variation in potato tuber yield in prince edward island," *Sustainability*, vol. 12, no. 12, p. 4937, 2020.

[3] M. St. Pierre, "Prince edward island leads the way in potato production," Statistics Canada, 02 2023. [Online]. Available: https://www150.statcan.gc.ca/n1/pub/96-325-x/2021001/article/00002-eng.htm

[4] "Area, production and farm value of potatoes," Statistics Canada, 07 2025. [Online]. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210035801

[5] M. W. Nasir and Z. Toth, "Effect of drought stress on potato production: A review," *Agronomy*, vol. 12, no. 3, p. 635, 2022.

[6] T. Van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Computers and electronics in agriculture*, vol. 177, p. 105709, 2020.

[7] E.-S. M. El-Kenawy, A. A. Alhussan, N. Khodadadi, S. Mirjalili, and M. M. Eid, "Predicting potato crop yield with machine learning and deep learning for sustainable agriculture," *Potato Research*, vol. 68, no. 1, pp. 759–792, 2025.

[8] A. Wolanin, G. Mateo-García, G. Camps-Valls, L. Gómez-Chova, M. Meroni, G. Duveiller, Y. Liangzhi, and L. Guanter, "Estimating and understanding crop yields with explainable deep learning in the indian wheat belt," *Environmental research letters*, vol. 15, no. 2, pp. 24–19, 2020.

[9] D. Tamayo-Vera, M. Mesbah, Y. Zhang, and X. Wang, "Advanced machine learning for regional potato yield prediction: analysis of essential drivers," *npj Sustainable Agriculture*, vol. 3, no. 1, p. 12, 2025.

[10] J. Kaddour, A. Lynch, Q. Liu, M. J. Kusner, and R. Silva, "Causal machine learning: A survey and open problems," *arXiv preprint arXiv:2206.15475*, 2022.

[11] F. Wang, J. Li, D. Peng, Q. Yi, X. Zhang, J. Zheng, and S. Chen, "Estimating soybean yields using causal inference and deep learning approaches with satellite remote sensing data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[12] P. Cui and S. Athey, "Stable learning establishes some common ground between causal inference and machine learning," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 110–115, 2022.

[13] V. Sitokonstantinou, E. Diaz, J. Cerda-Bautista, M. Piles, I. N. Athanasiadis, I. Tsoumas, and G. Camps-Valls, "Causal machine learning for sustainable agriculture," in *NeurIPS 2025 Workshop on CauScien: Uncovering Causality in Science*.

[14] A. Henningsen, G. Low, D. Wuepper, T. Dalhaus, H. Storm, D. Belay, and S. Hirsch, "Estimating causal effects with observational data: Guidelines for agricultural and applied economists," *arXiv preprint arXiv:2508.02310*, 2025.

[15] R. Moraffah, M. Karami, R. Guo, A. Raglin, and H. Liu, "Causal interpretability for machine learning-problems, methods and evaluation," *Acm Sigkdd Explorations Newsletter*, vol. 22, no. 1, pp. 18–33, 2020.

[16] I. Tsoumas, G. Giannarakis, V. Sitokonstantinou, A. Koukos, D. Loka, N. Bartsotas, C. Kontoes, and I. Athanasiadis, "Evaluating digital agriculture recommendations with causal inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, 2023, pp. 14 514–14 522.

[17] Y. Jiang, M. Ramsay, F. Meng, and T. Stetson, "Characterizing potato yield responses to water supply in atlantic canada's humid climate using historical yield and weather data: Implications for supplemental irrigation," *Agricultural Water Management*, vol. 255, p. 107047, 2021.

[18] R. Raymundo, S. Asseng, R. Robertson, A. Petsakos, G. Hoogenboom, R. Quiroz, G. Hareau, and J. Wolf, "Climate change impact on global potato production," *European Journal of Agronomy*, vol. 100, pp. 87–98, 2018.

[19] J. Pearl, *Causality*. Cambridge university press, 2009.

[20] B. Schölkopf, "Causality for machine learning," in *Probabilistic and causal inference: The works of Judea Pearl*, 2022, pp. 765–804.

[21] J. Peters, P. Bühlmann, and N. Meinshausen, "Causal inference by using invariant prediction: identification and confidence intervals," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 78, no. 5, pp. 947–1012, 2016.

[22] X. Xu, Q. Liu, Y. Liu, Y. Li, Y. Chen, T. Lei, Y. Kuzyakov, W. Zhang, J. Chen, and T. Ge, "Novel soil health assessment framework for legume-based rotation farmland by interpretable machine learning with causal inference," *Computers and Electronics in Agriculture*, vol. 239, p. 111011, 2025.

[23] P. Schwab, L. Linhardt, S. Bauer, J. M. Buhmann, and W. Karlen, "Learning counterfactual representations for estimating individual dose-response curves," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 04, 2020, pp. 5612–5619.

[24] K. Colangelo and Y.-Y. Lee, "Double debiased machine learning nonparametric inference with continuous treatments," *Journal of Business & Economic Statistics*, vol. 44, no. 1, pp. 67–79, 2026.

[25] J. D. Huling, N. Greifer, and G. Chen, "Independence weights for causal inference with continuous treatments," *Journal of the American Statistical Association*, vol. 119, no. 546, pp. 1657–1670, 2024.

[26] Z. Branson, E. H. Kennedy, S. Balakrishnan, and L. Wasserman, "Causal effect estimation after propensity score trimming with continuous treatments," *arXiv preprint arXiv:2309.00706*, 2023.

[27] L. Li, W. Shangguan, Y. Deng, J. Mao, J. Pan, N. Wei, H. Yuan, S. Zhang, Y. Zhang, and Y. Dai, "A causal inference model based on random forests to identify the effect of soil moisture on precipitation," *Journal of Hydrometeorology*, vol. 21, no. 5, pp. 1115–1131, 2020.

[28] S. Xue and G. Wu, "Causal inference of root zone soil moisture performance in drought," *Agricultural Water Management*, vol. 305, p. 109123, 2024.

[29] L. Zou, Y. Zha, Y. Diao, C. Tang, W. Gu, and D. Shao, "Coupling the causal inference and informer networks for short-term forecasting in irrigation water usage," *Water Resources Management*, vol. 37, no. 1, pp. 427–449, 2023.

[30] L. Umutoni and V. Samadi, "Application of machine learning approaches in supporting irrigation decision making: A review," *Agricultural Water Management*, vol. 294, p. 108710, 2024.

[31] E. Mwebaze, W. Okori, and J. A. Quinn, "Causal structure learning for famine prediction." in *AAAI Spring Symposium: Artificial Intelligence for Development*, 2010, pp. 61–66.

[32] H. Gao, L. Zhangzhong, W. Zheng, and G. Chen, "How can agricultural water production be promoted? a review on machine learning for irrigation," *Journal of Cleaner Production*, vol. 414, p. 137687, 2023.

[33] B. Zhang, A. A. Tsiatis, E. B. Laber, and M. Davidian, "A robust method for estimating optimal treatment regimes," *Biometrics*, vol. 68, no. 4, pp. 1010–1018, 2012.

[34] C. Liu, Y. Xu, Q. Liu, H. Zhu, and Y. Wang, "Application of machine learning based methods in exposure–response analysis," *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 49, no. 4, pp. 401–410, 2022.

[35] I. Bica, A. M. Alaa, C. Lambert, and M. Van Der Schaar, "From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges," *Clinical Pharmacology & Therapeutics*, vol. 109, no. 1, pp. 87–100, 2021.

[36] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, "Metalearners for estimating heterogeneous treatment effects using machine learning," *Proceedings of the national academy of sciences*, vol. 116, no. 10, pp. 4156–4165, 2019.

[37] K. Hirano, G. W. Imbens, and G. Ridder, "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, vol. 71, no. 4, pp. 1161–1189, 2003.

[38] D. W. Brown, T. J. Greene, M. D. Swartz, A. V. Wilkinson, and S. M. DeSantis, "Propensity score stratification methods for continuous treatments," *Statistics in medicine*, vol. 40, no. 5, pp. 1189–1203, 2021.

[39] T. Bahadori, E. T. Tchetgen, and D. Heckerman, "End-to-end balancing for causal continuous treatment-effect estimation," in *International conference on machine learning*. PMLR, 2022, pp. 1313–1326.

[40] M. L. Petersen, K. E. Porter, S. Gruber, Y. Wang, and M. J. Van Der Laan, "Diagnosing and responding to violations in the positivity assumption," *Statistical methods in medical research*, vol. 21, no. 1, pp. 31–54, 2012.

[41] E. A. Abioye, O. Hensel, T. J. Esau, O. Elijah, M. S. Z. Abidin, A. S. Ayobami, O. Yerima, and A. Nasirahmadi, "Precision irrigation management using machine learning and digital farming solutions," *AgriEngineering*, vol. 4, no. 1, pp. 70–103, 2022.

[42] "Canadian potato production 2025," *Statistics Canada*, 2025.

[43] "Canadian soil information service," Agriculture and Agri-Food Canada (AAFC), 2025. [Online]. Available: https://sis.agr.gc.ca/cansis

[44] T. J. Johansen, M. G. Thomsen, A.-K. Løes, and H. Riley, "Root development in potato and carrot crops–influences of soil compaction," *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, vol. 65, no. 2, pp. 182–192, 2015.

[45] J. Adriano Albuquerque, A. Da Costa, G. H. Merten, A. C. De Mattos E Avila, and G. Kirchhof, "Size of sand grains controls pore structure and water dynamics: Implications for water retention and hydraulic conductivity," *Land*, vol. 15, no. 5, p. 864, 2026.

[46] M. STALHAM, E. ALLEN, A. ROSENFELD, and F. HERRY, "Effects of soil compaction in potato."

[47] N. N. McKenzie, D. D. Jacquier, R. R. Isbell, and K. K. Brown, *Australian soils and landscapes: an illustrated compendium*. CSIRO publishing, 2004.

[48] D. Hillel, *Introduction to environmental soil physics*. Elsevier, 2003.

[49] R. G. Allen, L. S. Pereira, D. Raes, M. Smith *et al.*, "Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56," *Fao, Rome*, vol. 300, no. 9, p. D05109, 1998.

[50] E. G. Cutter, "Structure and development of the potato plant," in *The potato crop: the scientific basis for improvement*. Springer, 1978, pp. 70–152.

[51] P. Grassini, K. M. Eskridge, and K. G. Cassman, "Distinguishing between yield advances and yield plateaus in historical crop production trends," *Nature communications*, vol. 4, no. 1, p. 2918, 2013.

[52] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 2, p. e1449, 2022.

[53] P. Spirtes, C. N. Glymour, and R. Scheines, *Causation, prediction, and search*. MIT press, 2000.

[54] A. Sharma and E. Kiciman, "Dowhy: An end-to-end library for causal inference," *arXiv preprint arXiv:2011.04216*, 2020.

[55] M. M. Joffe, W. P. Yang, and H. I. Feldman, "Selective ignorability assumptions in causal inference," *The International Journal of Biostatistics*, vol. 6, no. 2, 2010.

[56] R. W. Schafer, "What is a savitzky-golay filter?[lecture notes]," *IEEE Signal processing magazine*, vol. 28, no. 4, pp. 111–117, 2011.

[57] D. H. Fleisher, D. J. Timlin, and V. R. Reddy, "Temperature influence on potato leaf and branch distribution and on canopy photosynthetic rate," *Agronomy Journal*, vol. 98, no. 6, pp. 1442–1452, 2006.

[58] C. Wagg, S. Hann, Y. Kupriyanovich, and S. Li, "Timing of short period water stress determines potato plant growth, yield and tuber quality," *Agricultural Water Management*, vol. 247, p. 106731, 2021.

[59] W. E. Marcílio and D. M. Eler, "From explanations to feature selection: assessing shap values as feature selection mechanism," in *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)*. Ieee, 2020, pp. 340–347.

[60] A. Karpatne, X. Jia, and V. Kumar, "Knowledge-guided machine learning: Current trends and future prospects," *arXiv preprint arXiv:2403.15989*, 2024.

[61] C. Ma, Z. Zhou, H. Liu, and D. Koslicki, "Kgml-xdtd: a knowledge graph–based machine learning framework for drug treatment prediction and mechanism description," *GigaScience*, vol. 12, p. giad057, 2023.

Appendix A

List of Variables in Metadata

Category	Feature	Abbreviation	Description
Crop	Yield	Yield (t/ac)	Marketable yield at postal-code level (t/ac).
	Lagged yield (1-year)	Lag Yield	Previous-year yield for the same postal code (t/ac).
	Moving average (3-year)	Moving Avg	Three-year lagged mean yield for the same postal code (t/ac).
Climate	GS water deficit/excess	GS Water Def Excess	Growing-season water balance (PRCP minus ETc) (mm; SS/MS/LS, GS).
	Acc. rainfall before planting	Acc Rain bf Plant	Pre-plant precipitation, Jan–Apr (mm).
	Acc. GDD before planting	Acc GDD bf Plant	Pre-plant GDD (base 7°C), Jan–Apr (°C-days).
	Mean wind speed	Mean Wind Spd	Seasonal mean wind speed (m/s; SS/MS/LS, GS).
	Minimum temperature	Min Temp	Seasonal mean daily minimum temperature (°C; SS/MS/LS, GS).
	Maximum temperature	Max Temp	Seasonal mean daily maximum temperature (°C; SS/MS/LS, GS).
	Vapor pressure deficit	VPD	Seasonal mean vapor pressure deficit (kPa; SS/MS/LS, GS).
	Radiation	Radiation	Total shortwave radiation over the growing season (MJ/m²).
	Precipitation	PRCP	Seasonal total precipitation (mm; SS/MS/LS, GS).
	Crop evapotranspiration	ETc	Seasonal total crop evapotranspiration (mm; SS/MS/LS, GS).
	Growing degree days	GDD7	Seasonal sum of GDD (base 7°C) after planting (°C-days; SS/MS/LS, GS).
Soil	Coarse fragments	COFRAG	Coarse fragments, 0–60 cm (%).
	Very fine sand	VFSAND	Very fine sand, 0–60 cm (%).
	Total sand	TSAND	Total sand, 0–60 cm (%).
	Total silt	TSILT	Total silt, 0–60 cm (%).
	Total clay	TCLAY	Total clay, 0–60 cm (%).
	Organic carbon	ORGCARB	Organic carbon, 0–60 cm (%).
	pH ($CaCl_2$)	PHCA	Soil pH ($CaCl_2$), 0–60 cm.
	pH (project)	PH2	Soil pH (project method / H_2O), 0–60 cm.
	Base saturation	BASES	Base saturation, 0–60 cm (%).
	Cation exchange capacity	CEC	Cation exchange capacity, 0–60 cm ($cmol_c/kg$).
	Saturated hydraulic conductivity	KSAT	Saturated hydraulic conductivity, 0–60 cm (cm/h).
	Water retention at 0 kPa	KP0	Volumetric water content at 0 kPa (saturation), 0–60 cm (%).
	Water retention at 33 kPa	KP33	Volumetric water content at 33 kPa (field capacity), 0–60 cm (%).
	Water retention at 1500 kPa	KP1500	Volumetric water content at 1500 kPa (wilting point), 0–60 cm (%).
	Bulk density	BD	Bulk density, 0–60 cm (g/cm³).
	Available water capacity	AWCmm	Available water capacity, 0–60 cm (mm).

Table 5: Summary of yield, climate, and soil variables.