

ACCEPTED MANUSCRIPT • OPEN ACCESS

Non-linear model-based sequential decision-making in agriculture

To cite this article before publication: Sakshi Arya *et al* 2026 *Environ. Res. Commun.* in press <https://doi.org/10.1088/2515-7620/ae489e>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2026 The Author(s). Published by IOP Publishing Ltd.



As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 4.0 licence, this Accepted Manuscript is available for reuse under a CC BY 4.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/4.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Nonlinear Model-Based Sequential Decision-Making in Agriculture

¹, Sakshi Arya^{*1}, and Wentao Lin^{†1,2}

¹Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve
University, Cleveland, OH 44106, USA

²CTrees, Pasadena, CA 91105, USA

Abstract

Agricultural decision-making faces a dual challenge: sustaining high yields to meet global food security needs while reducing the environmental impacts of input use, including fertilizer losses (e.g., nutrient runoff and leaching) and other agrochemical applications such as herbicides, insecticides, and fungicides. Nitrogen inputs are central to this tension. They are indispensable for crop growth yet also major drivers of greenhouse gas emissions, nutrient runoff, and escalating production costs. Addressing these intertwined pressures requires adaptive decision-support tools that are not only statistically principled but also economically sustainable and interpretable for practitioners.

We develop nonlinear model-based bandit algorithms as a framework for adaptive fertilizer management under uncertainty. Building on classical mechanistic yield-response models including Mitscherlich, Michaelis-Menten, quadratic plateau, and logistic functions, our approach links algorithmic exploration-exploitation strategies directly to interpretable biological processes such as maximum yield and nutrient efficiency. This grounding makes recommendations transparent for practitioners while supporting cost-effective and sustainable input use. Methodologically, we establish regret and sample complexity results for the well-specified nonlinear case, examine robustness under misspecification, and evaluate the proposed methods through extensive profit-oriented simulations and an offline replay case study on publicly available multi-site corn nitrogen field trials from the U.S. Midwest. Overall, the results show that incorporating biologically meaningful mechanistic structure enables faster learning and higher profit as evidence accumulates, with flexible nonparametric baselines providing a competitive alternative in pooled and heterogeneous settings. Our findings illustrate how interpretable, uncertainty-aware sequential decision rules can support economically sustainable fertilizer recommendations and contribute to more efficient agricultural input use.

Keywords: Sequential decision-making, agricultural statistics, nonlinear models, multi-armed bandits, resource optimization, regret

1 Introduction

Identifying the most effective crop management practice for a given site is a recurring challenge in agriculture. While field trials and expert recommendations provide useful starting points, the

^{*}Corresponding author. Email: sxa1351@case.edu

[†]Email: wentao.lin223@gmail.com (Part of this work was done when the author was a student at Case Western Reserve University.)

1
2
3 actual performance of management strategies such as different fertilizer rates, planting densities,
4 or irrigation schedules can be highly variable due to differences in soil properties, previous man-
5 agement, and other biological or environmental factors (Hochman and Carberry, 2011; Tilman
6 et al., 2002). As a result, farmers and researchers must repeatedly decide which management ac-
7 tions to implement each season, often with limited prior information about their true effects. This
8 problem exemplifies *sequential decision-making under uncertainty*, where the goal is to efficiently
9 learn and optimize management choices over time in a way that supports both productivity and
10 sustainability.

11
12 In the statistical and machine learning literature, such problems are naturally framed as *multi-*
13 *armed bandit* problems, dating back to Robbins (1952). They have been widely applied in areas
14 such as precision medicine (Lu et al., 2021), recommendation systems (Li et al., 2010), and, in-
15 creasingly, agriculture (Gautron et al., 2022). The term originates from the analogy of a gambler
16 choosing between several slot machines (“one-armed bandits”), each with unknown payout rates.
17 At each round, the decision-maker selects an *action/arm* (e.g., a specific management practice)
18 and observes its outcome (e.g., crop yield or profit). The central challenge is to balance *exploration*
19 (trying new or uncertain options to learn about their effects) with *exploitation* (repeating choices
20 that have performed well so far). Bandit algorithms provide principled, data-driven strategies to
21 manage this tradeoff, making them well-suited to resource optimization tasks in agriculture and
22 environmental management. A standard way to evaluate such algorithms is through minimiz-
23 ing *regret* (see Table 1 for formal definition), which measures the cumulative loss incurred while
24 learning compared to an ideal benchmark that always selects the best fixed action in hindsight.

25
26 Most existing bandit algorithms rely on linear/GLM structure or generic black-box models, which
27 can miss the mechanistic nonlinear dose-response patterns common in agronomy. We develop a
28 family of bandit algorithms that embeds agronomy-standard nonlinear response models directly in
29 the decision loop, yielding interpretable recommendations and improved sample efficiency in the
30 small-sample regimes typical of field experimentation, where decisions carry meaningful economic
31 and environmental stakes.

32
33 Nonlinear models provide parsimonious and interpretable representations of domain-specific pro-
34 cesses. Unlike generic polynomial or linear models, their parameters often map directly to biological
35 quantities, such as maximum yield potential or nutrient efficiency, making the models both sci-
36 entifically meaningful and statistically efficient. In agriculture, such models (Miguez et al., 2018)
37 capture complex responses like crop yield to fertilizer, temperature effects on growth, or pest dy-
38 namics, where the relationship follows saturating, logistic, or exponential patterns rather than a
39 simple linear trend. Their adoption is motivated not only by statistical accuracy but also by the
40 need for transparent recommendations in sustainable resource management

41 42 43 44 45 46 **1.1 Background and related work**

47 Multi-armed bandit algorithms (Berry and Fristedt, 1985; Lattimore and Szepesvári, 2018) pro-
48 vide a core statistical framework for sequential decision-making under uncertainty, where the key
49 challenge is to balance exploration and exploitation. In the original bandit formulation, each arm
50 was associated with an unknown fixed reward distribution, and repeated pulls of the same arm
51 were assumed to yield independent and identically distributed (i.i.d.) rewards. Subsequent work
52 extended this framework to incorporate contextual information, leading to the development of con-
53 textual bandit models (Slivkins, 2014; Chu et al., 2011). In these settings, regression models such
54 as linear (Abbasi-Yadkori et al., 2011), generalized linear models (GLMs) (Li et al., 2010; Filippi
55

1
2
3 et al., 2010), and more flexible nonparametric approaches, including kernels, nearest neighbors,
4 and neural networks (Yang and Zhu, 2002; Rigollet and Zeevi, 2010; Srinivas et al., 2010; Valko
5 et al., 2013), enable adaptive decision-making tailored to evolving arm/covariate information.
6

7 Classical nonlinear regression models, such as Mitscherlich (Dhanoa et al., 2022), Michaelis-Menten
8 (Lopez et al., 2000), quadratic plateau models (Bélanger et al., 2000), and logistic response func-
9 tions (Sepaskhah et al., 2011), are routinely used to capture crop yield response to fertilizer and
10 other management variables. These models offer interpretable parameters tied to underlying bio-
11 logical processes, which is important for transparent and actionable recommendations (De Bruin
12 and Pedersen, 2008; Miguez et al., 2018).
13

14 Machine learning approaches, including random forests, neural networks, and regression trees, have
15 become increasingly popular for crop yield prediction and resource optimization (Shahhosseini
16 et al., 2019; Khaki et al., 2020; Khairunniza-Bejo et al., 2014; Javed and Murad, 2024). While
17 powerful in large-data settings, these methods are typically supervised and static, lacking the
18 sequential and adaptive capabilities needed for decision-making under uncertainty. Concerns also
19 remain about their interpretability and suitability for low-data environments (Dobermann et al.,
20 2022).
21
22

23 Within agriculture, adaptive experimentation and real-time learning strategies have emerged as
24 promising approaches for site-specific management and resource efficiency (Gautron et al., 2022).
25 Recent work has explored the use of bandit algorithms (Saikai et al., 2020, 2018; Huang, 2025) for
26 adaptive plot selection and management optimization, but most existing studies focus on linear
27 models or generic black-box methods, seldom leveraging domain knowledge of nonlinear response
28 behavior, especially valuable for smallholders and in data-scarce scenarios (Smith, 2018). Along-
29 side bandit and reinforcement-learning approaches, a growing precision-agriculture literature uses
30 Gaussian-process surrogate modeling, Bayesian optimization, and Bayesian experimental design to
31 adaptively select treatment levels and sampling schemes in on-farm trials, often with an explicit fo-
32 cus on estimating the economic optimum nitrogen rate (EONR). For example, Ngómbe and Brorsen
33 (2022) develop Bayesian optimal dynamic sampling procedures for on-farm experimentation, and
34 Matavel et al. (2025) propose Bayesian-optimized experimental designs for EONR estimation using
35 a model-averaging approach. Related work develops fully Bayesian economically optimal designs
36 in spatial and multi-year settings (Poursina and Brorsen, 2024), uses Gaussian-process modeling
37 across multiple on-farm experiments to quantify uncertainty and profitability of fertilizer strategies
38 (Mia et al., 2024), and advances multi-environment evaluation of nitrogen recommendation tools,
39 highlighting the need for decision rules that can update across seasons (Abdipourchenarestansofla
40 and Piepho, 2025; Ransom et al., 2020). Bayesian optimization has also been used to automate
41 calibration of mechanistic crop models, illustrating its role in decision-support pipelines (Akhav-
42 izadegan et al., 2021). Our work is complementary to these Bayesian optimization and design
43 approaches. While they primarily target efficient experimentation and estimation (e.g., learn-
44 ing EONR with minimal trials) via a learned surrogate and an acquisition rule, we study bandit
45 decision rules that explicitly balance exploration and exploitation to optimize *cumulative* perfor-
46 mance (profit and/or yield) over repeated decisions. Moreover, surrogate/acquisition formulations
47 can be less directly interpretable for agronomic decision-making, whereas our policies are built
48 around mechanistic dose-response models whose parameters map to familiar agronomic quanti-
49 ties (e.g., plateau level, response rate), making the learned recommendations more transparent in
50 data-limited settings.
51
52
53
54

55 It is well-known that model-based bandit and reinforcement learning approaches can be far more
56 sample-efficient than model-free methods when the underlying reward mechanisms are known (Os-
57
58
59
60

band and Van Roy, 2014; Sun et al., 2019). However, such approaches have received limited attention in agricultural decision-making from a precision-agriculture and sustainability perspective, where established nonlinear models are often available but underutilized. Our work addresses this gap by integrating domain-driven nonlinear models directly into the bandit framework. This enables adaptive, interpretable, and sample-efficient resource management, advancing the practical and theoretical toolkit for sustainable and cost-effective fertilizer use.

Our primary objective is to adapt nitrogen-rate recommendations over repeated seasons by balancing exploration and exploitation to maximize overall profit while learning the profit-maximizing rate (EONR / maximum return to nitrogen) over a feasible fertilizer-rate grid.

Our Contributions. This paper makes the following contributions:

- **Mechanistic nonlinear bandits for fertilizer-rate decisions:** We formulate nitrogen-rate selection as a sequential decision problem in which the mean reward is modeled using agronomy-standard mechanistic nonlinear dose–response families (Mitscherlich, quadratic plateau, Michaelis–Menten, and logistic). To our knowledge, this is the first bandit-based framework in agronomic decision-making that provides a unified, interpretable template for mechanistic nonlinear response models rather than restricting to linear/GLM rewards or fully model-free black-box learners, and it can be adapted to different agronomic objectives by redefining the reward (e.g., profit, yield, or other management utilities).
- **Nonlinear model-based ϵ -greedy, nonlinear-UCB, and ViOlin baselines:** We present a unified framework and pseudocode templates for three nonlinear model-based strategies, ϵ -greedy, UCB, and ViOlin, that can be instantiated with the above mechanistic response models, providing an interpretable alternative to linear and model-free baselines for data-limited agronomic decision-making.
- **Profit-oriented empirical evaluation in simulated and real trials:** Under the profit objective (learning the EONR), we conduct an extensive simulation study under both well-specified and misspecified regimes, and we additionally provide a reproducible real-data case study using multi-site corn nitrogen trials from Ransom et al. (2021) via an offline replay protocol. These experiments quantify when mechanistic nonlinear modeling yields large gains and how performance degrades under model mismatch, relative to linear methods (e.g., LinUCB) and fully model-free methods (e.g., kNN-UCB), particularly in low-sample regimes.
- **Theory-to-practice connection via available complexity and sample-efficiency results:** We connect the nonlinear model classes considered here to existing sequential learning theory by presenting a sequential complexity bound for bounded function classes (with an illustrative visualization for the four nonlinear families studied) and by summarizing available sample-efficiency results for curvature-guided methods such as ViOlin from Dong et al. (2021), discussing its implications for data-limited agricultural experimentation.

We begin by setting up the problem in Section 2. In Section 3, we lay out the proposed framework for nonlinear model-based bandits, in which we study three model-based algorithms: ϵ -greedy, Upper Confidence Bound (UCB), and the ViOlin algorithm. Section 4 presents known results for parametric bandits in both well-specified and misspecified settings, and derives sample complexity bounds for the specific nonlinear models considered by the ViOlin algorithm. In Section 5, we then conduct extensive simulation experiments emulating real-world yield production and fertilizer

optimization scenarios, with the goal of resource-efficient and cost-effective management in sample-limited settings. Then, in Section 6 we conduct a reproducible real-data case study based on multi-site corn nitrogen trials in the US midwest to illustrate the advantage of the proposed methodology. Finally, Section 7 concludes the paper.

2 Problem Setup: Adaptive Decisions with Nonlinear Models

Formally, we consider a sequential decision-making scenario with a time horizon of T rounds. Each round can represent, for example, a growing season or field trial, where decisions are made at planting and outcomes observed at harvest. At each round $t = 1, \dots, T$, the decision-maker (agent) chooses an action (input level) $x_t \in \mathcal{X}$ from a feasible set \mathcal{X} , such as nitrogen application rates. After choosing action x_t , the agent observes a noisy reward y_t , representing yield or profit, modeled as:

$$y_t = f(x_t; \beta^*) + \eta_t, \quad (1)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$ is a known, structured nonlinear function parameterized by an unknown vector $\beta^* \in \mathbb{R}^p$, and η_t is a mean-zero noise term. We assume η_t is *sub-Gaussian*, i.e., there exists $\sigma > 0$ such that for all $\lambda \in \mathbb{R}$,

$$\mathbb{E} \left[e^{\lambda \eta_t} \right] \leq \exp \left(\frac{\lambda^2 \sigma^2}{2} \right), \quad (2)$$

where $\mathbb{E}(\cdot)$ denotes the expected value with respect to the random noise. This assumption reflects the variability of agricultural experiments driven by weather, soil heterogeneity, and biological processes, while ruling out extreme outliers that would make learning impractical.

The decision-maker seeks a policy/algorithm π that selects actions based on historical observations to maximize cumulative expected reward. Performance can equivalently be measured through the cumulative regret, which quantifies the total loss relative to always applying the best treatment:

$$R_T(\pi) = \sum_{t=1}^T \{f(x^*; \beta^*) - f(x_t; \beta^*)\}, \quad (3)$$

where $x^* = \arg \max_{x \in \mathcal{X}} f(x; \beta^*)$ denotes the optimal treatment/arm. In agricultural terms, regret measures the cumulative yield or profit lost due to not choosing the best possible input at each season. The objective is to devise algorithms achieving *sublinear regret*, meaning $R_T(\pi) = o_P(T)$, so that the average regret per round diminishes over time. Such a property guarantees that the policy rapidly identifies near-optimal input levels, minimizing wasted resources and experimental costs. Other performance metrics more suitable for risk-averse decision-making have also been proposed, such as the Conditional Value-at-Risk (CVaR) of the regret, which emphasizes control of the worst-case outcomes in the tail of the regret distribution (Baudry et al., 2021; Wang et al., 2023).

Remark 1 (Scope). *For clarity, we study a stationary setting where the underlying mean response does not change across rounds, so the optimal benchmark rate x^* is fixed. Accounting for season-to-season shifts (e.g., weather or soil changes) would require a nonstationary model and is left for future work.*

2.1 From yield to profit: an economic objective.

In fertilizer management, maximizing yield alone is not necessarily the right decision objective. Most yield–nitrogen response curves exhibit diminishing returns: yield increases with fertilizer up to a point and then plateaus. Because fertilizer is a costly input, applying nitrogen beyond the economically efficient range can increase cost without proportional yield gains. For this reason, we focus on *profit* (net return) rather than yield as the primary optimization target.

Let p_y denote the price per unit yield (e.g., \$/bushel) and let p_x denote the price per unit nitrogen (e.g., \$/lb N). Suppose the mean yield response is modeled by a parametric function $f(x; \theta)$, where x is the nitrogen rate and θ denotes agronomic parameters. We define the profit function

$$\Pi(x) = p_y f(x; \theta) - p_x x. \quad (4)$$

The economically optimal fertilizer rate is then

$$x^*(\theta) = \arg \max_{x \in \mathcal{X}} \Pi(x) = \arg \max_{x \in \mathcal{X}} [p_y f(x; \theta) - p_x x], \quad (5)$$

where \mathcal{X} is the feasible set of fertilizer rates (e.g., a grid from 0 to 250 lb N/ac). Depending on the form of f , this maximization can be carried out in closed form or numerically. In our experiments, we use closed-form expressions for x^* for the yield-response models considered (summarized in Table 2), which makes clear how the optimal decision depends jointly on agronomic parameters θ and economic parameters (p_y, p_x) that must be learned from data. Accordingly, we define (expected) profit regret as $\Pi(x^*) - \Pi(x_t)$ and evaluate cumulative regret $\sum_{t=1}^T \{\Pi(x^*) - \Pi(x_t)\}$, where x_t is the fertilizer rate selected by the algorithm at time t . For the nonlinear yield functions studied in our simulations, closed-form expressions for x^* are summarized in Table 2.

Why bandits? Although $x^*(\theta)$ admits closed-form expressions (Table 2), θ is unknown and must be learned online from limited data; the bandit algorithms couple plug-in optimization with explicit exploration to avoid premature convergence under early estimation error.

To illustrate the framework, we focus on fertilizer-yield relationships and investigate four classical nonlinear models widely used in agricultural research:

1. **Mitscherlich Model:** $f(x; A, b, d) = d + A(1 - e^{-bx})$, capturing saturating yield response to fertilizer inputs.
2. **Michaelis-Menten Model:** $f(x; a, b, d) = d + \frac{ax}{b+x}$, describing nutrient uptake or growth responses with diminishing returns.
3. **Quadratic Plateau Model:**

$$f(x; a, b, c, x_0) = \begin{cases} a + bx + cx^2 & \text{if } x \leq x_0 \\ a + bx_0 + cx_0^2 & \text{if } x > x_0 \end{cases},$$

modeling responses that rise and then stabilize at a plateau, common in fertilizer trials.

4. **Logistic Dose-Response Model:** $f(x; A, B, C, d) = d + \frac{A}{1 + \exp(-B(x-C))}$, often used to capture threshold or inflection-point behavior.

These models are parsimonious, interpretable, and directly tied to biological processes. For example, in the Mitscherlich model, A represents the maximum additional yield achievable with

Table 1: Key concepts and notation for sequential fertilizer-rate decision-making.

Term	Meaning in this paper
Round ($t = 1, \dots, T$)	One decision opportunity (e.g., a season, site-year, or trial iteration) in which a nitrogen rate is selected and an outcome is observed.
Action / arm ($x_t \in \mathcal{X}$)	The nitrogen fertilizer rate applied at round t ; \mathcal{X} is a feasible set (often a discrete grid).
Reward (y_t)	Observed outcome from choosing x_t (yield or profit, depending on the objective).
Policy / algorithm (π)	The rule the decision-maker follows each round: it looks at all past fertilizer choices and outcomes $\{(x_i, y_i)\}_{i=1}^{t-1}$ and then recommends the next nitrogen rate x_t .
Bandit (partial) feedback	Only the outcome corresponding to the chosen rate x_t is observed at each round (not outcomes for all rates).
True mean reward ($f(x; \beta^*)$)	Expected outcome at rate x under the unknown true parameter β^* .
Estimated mean reward ($f(x; \hat{\beta}_t)$)	Plug-in prediction used by the algorithm at round t based on data collected so far.
Exploration vs. exploitation	Trying uncertain rates to learn (exploration) versus choosing the currently best-performing rate (exploitation).
Economic objective ($\Pi(x)$)	Profit at rate x , typically $\Pi(x) = p_y Y(x) - p_x x$, with crop price p_y and input cost p_x .
Optimal rate (x^*)	The best fixed rate in \mathcal{X} under the objective (e.g., the profit-maximizing rate / EONR).
Warm start (n_0)	Initial rounds of forced exploration (uniformly random choices) to stabilize early model fitting before using model-based rules.
Per-round regret (r_t)	Loss relative to the benchmark: $r_t = \Pi(x^*) - \Pi(x_t)$ in simulations; in offline replay, x^* is the within-round oracle based on observed treatment means.
Cumulative regret (R_t) and average regret (R_t/t)	$R_t = \sum_{s=1}^t r_s$ summarizes total loss; R_t/t summarizes average loss per decision (used for “decreasing” curves).

fertilizer, b reflects the efficiency of nutrient uptake, and d is the baseline yield. For example, a larger value of b indicates that yield saturates more rapidly with increasing input, whereas a smaller b corresponds to a more gradual response. Such parameters are routinely used in agronomic practice and make the resulting recommendations not only statistically grounded but also transparent and actionable.

Although we emphasize nitrogen fertilizer management as a motivating example, the framework generalizes to a wide range of agricultural and environmental problems where nonlinear dose-response relationships are well established, including irrigation efficiency, biomass accumulation, and nutrient loss in response to rainfall (Miguez et al., 2018). We provide detailed parameteriza-

tions and simulation examples in Section 5.

3 Methods: Algorithms for nonlinear model-based bandits

In agricultural trials, each season offers only limited opportunities for experimentation. Algorithms must therefore use past data efficiently while remaining robust to uncertainty. In this section, we describe a family of strategies for nonlinear model-based bandit problems, where the expected reward is modeled as a nonlinear function of the chosen input. We first provide a general algorithmic template, followed by detailed descriptions of three model-based strategies, namely, ϵ -greedy, Upper Confidence Bound (UCB), and `Vi01in`, adapted to classical agronomic response models. Note that all the three algorithms have been studied in various bandit problems before (Auer and Ortner, 2010; Arya and Yang, 2020; Dong et al., 2021), here we specialize them to mechanistic nonlinear settings.

3.1 General framework

Recall, $f(x; \theta)$ denotes a known nonlinear model (e.g., Michaelis-Menten, logistic, quadratic) parameterized by θ , where $x \in \mathcal{X}$ is the action (e.g., N-rate) and the expected reward is $\mathbb{E}[Y \mid x] = f(x; \theta)$. At each time t , the agent selects an action x_t , then receives a reward y_t .

Algorithm 1 General Nonlinear Model-based Bandit Framework

- 1: **Input:** Action set \mathcal{X} , nonlinear model class $f(\cdot; \theta)$, time horizon: T
 - 2: **Initialize:** Pull each arm at least once and receive corresponding rewards until round t_0
 - 3: **for** $t = t_0 + 1$ to T **do**
 - 4: Estimate model parameters $\hat{\theta}_t$ using past data $\{(x_s, y_s) : s < t\}$
 - 5: Choose action x_t using a strategy/policy π utilizing $f(x; \hat{\theta}_t)$
 - 6: Observe reward y_t
 - 7: Update dataset with (x_t, y_t)
 - 8: **end for**
-

Illustrative example: To make Algorithm 1 concrete, consider a small pilot dataset with 5 nitrogen rates (e.g., $\mathcal{X} = \{0, 50, 100, 150, 200\}$) and 2 independent replicated plots per rate, giving $t_0 = 10$ observed pairs $\{(x_i, y_i)\}_{i=1}^{10}$. At the first adaptive decision round (i.e., $t = 11$), Line 4 fits the nonlinear model using *all* available observations, so the nonlinear least-squares objective uses 10 data points and returns $\hat{\theta}_{11}$. Line 5 then selects the next rate x_{11} from \mathcal{X} using the chosen policy/bandit algorithm applied to $f(\cdot; \hat{\theta}_{11})$. After observing y_{11} in Line 6, Line 7 appends (x_{11}, y_{11}) , so the next update uses 11 observations. In general, at round t , $\hat{\theta}_t$ is estimated from the full history $\{(x_s, y_s)\}_{s=1}^{t-1}$ (i.e., $t - 1$ observations), after which the bandit policy (algorithm) π selects the next action x_t based on $f(\cdot; \hat{\theta}_t)$.

In our implementation, $\hat{\theta}_t$ is computed via nonlinear least squares (e.g., using `curve_fit` in `scipy.optimize`) by minimizing a squared-error objective of the form

$$\hat{\theta}_t = \arg \min_{\theta} \sum_{s < t} (y_s - f(x_s; \theta))^2.$$

After observing y_t , the pair (x_t, y_t) is appended to the dataset and used in the next update.

Remark 2 (Bandit (partial) feedback). *Unlike classical supervised learning or factorial field experiments that provide outcomes for every treatment, bandit settings reveal only the reward for the action actually chosen at each round. This partial feedback captures the reality of sequential agronomic decision-making where only tested nitrogen rates yield observable outcomes. As a consequence, even when a closed-form optimizer $x^*(\theta)$ exists for a parametric response family, the parameters θ cannot be learned reliably without deliberate exploration, since unchosen rates provide no counterfactual information.*

Next, we describe three model-based algorithms that follow the general structure of Algorithm 1 but differ in how they balance the exploration-exploitation trade-off.

3.2 Epsilon-Greedy

The ϵ -greedy algorithm (Algorithm 2) (Langford and Zhang, 2008; Yang and Zhu, 2002) is one of the simplest approaches yet effective for balancing exploration and exploitation. At each round, the current nonlinear model is fitted using past observations, and the input that maximizes predicted profit is selected with high probability $1 - \epsilon_t$ (*exploitation*). With small probability ϵ_t , the algorithm explores by selecting a random alternative from the feasible input grid (*exploration*). This mechanism ensures continued learning, preventing the algorithm from becoming stuck on suboptimal choices. In agricultural contexts, this method reflects the intuitive practice of trying occasional alternative fertilizer rates even when one rate appears best.

Several variants of ϵ -greedy exist; the most popular is the *annealed* or *decaying* ϵ -greedy, where the exploration probability ϵ_t is a non-increasing sequence that tends to zero as $t \rightarrow \infty$. The intuition is that, as more data are collected and the model becomes more certain, the algorithm increasingly favors exploitation over exploration. Note that the user specifies an exploration *schedule* $\{\epsilon_t\}_{t \geq 1}$ (typically through an initial value and a decay rule), which controls the exploration–exploitation trade-off and directly affects cumulative regret.

Epsilon-greedy is simple, robust, and easy to implement. It is particularly useful when the model is misspecified or when simple, interpretable algorithms are desired.

Algorithm 2 Nonlinear model-based Epsilon-Greedy

- 1: **Input:** Action set \mathcal{X} , exploration schedule $\{\epsilon_t\}_{t \geq 1}$ (e.g., $\epsilon_t = \min\{1, \epsilon_0 t^{-\gamma}\}$), non-linear model class $f(\cdot; \theta)$, horizon T
 - 2: **Initialize:** Random pull arms and receive corresponding rewards until round t_0
 - 3: **for** $t = t_0 + 1$ to T **do**
 - 4: Estimate $\hat{\theta}_t$ from previous data: $\{x_1, y_1, \dots, x_{t-1}, y_{t-1}\}$
 - 5: Select action x_t according to:

$$x_t = \begin{cases} \text{sample uniformly from } \mathcal{X} & \text{with probability } \epsilon_t \\ \arg \max_{x \in \mathcal{X}} f(x; \hat{\theta}_t) & \text{with probability } 1 - \epsilon_t \end{cases}$$
 - 6: Observe y_t , update data
 - 7: **end for**
-

If \mathcal{X} is a finite set of K elements, “sample uniformly from \mathcal{X} ” means selecting an element at random with equal probability, so that each arm is chosen with probability $1/K$. In the ϵ -greedy algorithm,

during the exploration step, each arm is selected with probability ϵ/K . If \mathcal{X} is a continuous interval or region, this means sampling from the continuous uniform distribution over \mathcal{X} , i.e., $x_t \sim \text{Unif}(\mathcal{X})$.

3.3 Nonlinear model-based UCB

Overview: The UCB (Upper Confidence Bound) algorithm (Auer and Ortner, 2010; Chu et al., 2011; Zhou et al., 2020) embodies the principle of “optimism in the face of uncertainty” by selecting actions that maximize an upper confidence bound on the expected reward. Here, the nonlinear model is repeatedly fitted, and for each input level the algorithm constructs an upper confidence bound on the predicted reward. The next action is chosen to maximize this upper bound. This mechanism naturally prioritizes inputs that are either promising in terms of mean yield or remain highly uncertain, thereby encouraging exploration in scientifically justified directions. Applied to fertilizer management, UCB corresponds to testing input levels where either predicted yield is high or parameter uncertainty remains large.

In model-based nonlinear regression, the uncertainty $\text{Unc}_t(x)$ is often based on the variability of the estimated reward due to finite data, as quantified via the standard error of the model prediction. Specifically, the uncertainty term is given by

$$\text{Unc}_t(x) = \sqrt{\nabla_{\theta} f(x; \hat{\theta}_t)^{\top} \widehat{\text{Cov}}(\hat{\theta}_t) \nabla_{\theta} f(x; \hat{\theta}_t)}, \quad (6)$$

where $\nabla_{\theta} f(x; \hat{\theta}_t)$ denotes the gradient of the reward model with respect to the parameters, evaluated at the current estimate $\hat{\theta}_t$, and $\widehat{\text{Cov}}(\hat{\theta}_t)$ is the estimated covariance matrix of $\hat{\theta}_t$. We emphasize that $\text{Unc}_t(x)$ is a first-order error-propagation (delta-method) proxy and is not claimed to be a certified nonasymptotic confidence radius. This is nonetheless appropriate for our algorithmic use: the UCB rule only needs a data-adaptive ranking of uncertainty across candidate rates to avoid premature commitment when T is small. In early rounds, the nonlinear fit is typically weakly identified and $\widehat{\text{Cov}}(\hat{\theta}_t)$ is larger (and can be ill-conditioned), which increases $\text{Unc}_t(x)$ and naturally promotes additional exploration. To stabilize this regime we (i) warm-start with uniform exploration for the first n_0 observations and (ii) compute $\widehat{\text{Cov}}(\hat{\theta}_t)$ using ridge/regularized inversion of the observed information, preventing numerical degeneracy in very small samples. We use (6) as a practical and interpretable exploration score for data-limited agronomic settings, and evaluate its behavior empirically in our simulations and offline real-data replay experiments.

The UCB action selection rule then takes the form

$$x_t = \arg \max_{x \in \mathcal{X}} f(x; \hat{\theta}_t) + \alpha \cdot \text{Unc}_t(x) \quad (7)$$

where $\alpha > 0$ is a user-specified confidence parameter that controls the degree of exploration.

UCB algorithms provide strong theoretical guarantees for balancing exploration and exploitation, and are particularly effective when reliable, model-based estimates of prediction uncertainty are available (see, e.g., (Zhou et al., 2020)). It is particularly advantageous in settings where data collection is expensive or risky, as it prioritizes actions that could yield high rewards or that have not been thoroughly explored. Model-based UCB (Algorithm 6) thus combines statistical rigor with the agronomic interpretability of nonlinear functions, making it a strong candidate for adaptive on-farm experimentation.

Algorithm 3 Nonlinear model-based UCB

1: **Input:** Action set \mathcal{X} , confidence parameter $\alpha > 0$
2: **for** $t = 1$ to T **do**
3: Estimate $\hat{\theta}_t$ from the previous data
4: **for** each $x \in \mathcal{X}$ **do**
5: Compute predicted reward $f(x, z_t; \hat{\theta}_t)$ and uncertainty $\text{Unc}_t(x)$ as defined in (6)
6: Calculate $\text{UCB}_t(x) = f(x, z_t; \hat{\theta}_t) + \alpha \cdot \text{Unc}_t(x)$
7: **end for**
8: Choose $x_t = \arg \max_{x \in \mathcal{X}} \text{UCB}_t(x)$
9: Observe y_t , update data
10: **end for**

3.4 Vi0lin (Virtual Ascent with Online Model Learner)

The Vi0lin algorithm (Algorithm 4) (Dong et al., 2021) is a model-based bandit strategy designed for efficient learning in nonlinear settings. Unlike ϵ -greedy, which explores actions uniformly at random, or UCB, which prioritizes actions with high uncertainty, Vi0lin is a greedy method that leverages both the current model estimate and local geometric information about the reward surface (i.e., gradient and curvature). At each round, the algorithm selects the action predicted to be best by the current estimated model, while using local geometric information (slope and curvature) of the *fitted* reward surface to guide the search. In our implementation, this geometry enters through the action-selection score in Algorithm 4, and model parameters are updated from accumulated noisy rewards via standard parametric fitting. This geometry-guided greedy strategy is motivated by the ViOlin framework of Dong et al. (2021), which provides sample-efficiency guarantees when local gradient/curvature (Hessian) information about the reward function can be leveraged; here we use analytic derivatives of the fitted mechanistic model as a practical proxy under noisy observations.

Algorithm 4 Vi0lin: Virtual Ascent with Online Model Learner (Dong et al., 2021)

1: **Input:** Model class $\mathcal{F} = \{f(\cdot; \theta) : \theta \in \Theta\}$, action set \mathcal{X} , initial guess $\hat{\theta}_1$, learner/estimator \mathcal{O} , total rounds T , curvature weights $\kappa_1, \kappa_2 \geq 0$, minimum fit size m
2: **for** $t = 1$ to T **do**
3: **if** $t \leq m$ **then**
4: Choose x_t uniformly at random from \mathcal{X}
5: Observe noisy reward y_t
6: **else**
7: Fit/update $\hat{\theta}_t$ using $\{(x_s, y_s)\}_{s=1}^{t-1}$ via \mathcal{O}
8: **Action selection:** for each $x \in \mathcal{X}$ compute

$$\hat{\mu}_t(x) = f(x; \hat{\theta}_t), \quad \hat{g}_t(x) = \partial_x f(x; \hat{\theta}_t), \quad \hat{H}_t(x) = \partial_{xx}^2 f(x; \hat{\theta}_t),$$

and choose

$$x_t \in \arg \max_{x \in \mathcal{X}} \left\{ \hat{\mu}_t(x) + \kappa_1 |\hat{g}_t(x)| + \kappa_2 |\hat{H}_t(x)| \right\}.$$

9: Observe y_t , update data.
10: **end if**
11: **end for**

Note that, in our experiments, slope and curvature terms are computed from the fitted parametric model $f(\cdot; \hat{\theta}_t)$ (analytic derivatives) and used to guide action selection.

ViOlin is particularly attractive in data-limited settings (e.g., smallholder or resource-constrained agricultural experiments), where minimizing the number of field trials is crucial. **ViOlin**'s greedy exploration is guided by the model fit and its local geometry, without explicit randomized or uncertainty-driven action selection. This enables fast convergence to (local) optima when the model class matches the true reward structure and curvature information is reliable. In low-dimensional agronomic models with smooth response surfaces, **ViOlin** performs on par with simpler strategies such as ϵ -greedy and UCB. However, as model complexity (e.g., multi-modal response curves) or the dimensionality of \mathcal{X} increases, geometry-guided strategies such as **ViOlin** can be especially sample-efficient when the model is well specified and curvature information is reliable. This makes **ViOlin** well suited for adaptive experimentation in limited data regimes, such as on-farm field trials or precision agriculture for smallholder contexts, where each data point is costly to obtain.

Summary of algorithm differences:

ϵ -greedy explores uniformly at random, UCB targets actions with high uncertainty (optimism-based exploration), and **ViOlin** is greedy with geometry-guided action selection, using slope/curvature information from the fitted model to promote sample-efficient learning in nonlinear settings.

To illustrate how these algorithms operate and compare in practice, we include a simple, step-by-step example using the Mitscherlich model over three or four rounds in Section 5 of the Appendix. This concrete illustration should help readers, especially those less familiar with bandit algorithms, understand the exploration-exploitation tradeoff.

While the primary emphasis of this work is not on theoretical development, we present and discuss several well-known theoretical results and highlight their implications for the nonlinear model-based bandit framework.

4 Theoretical guarantees for nonlinear bandit algorithms

In this section, we characterize the regret and sample-complexity guarantees of sequential algorithms under increasingly flexible reward models, beginning with classical linear formulations, extending to nonparametric function classes, then focusing on nonlinear parametric (mechanistic) models as used in agronomy, and finally accounting for model misspecification. While the general theory applies broadly, our emphasis is on sample efficiency and robustness in low-data agricultural regimes, where each field trial is costly and decision errors translate to both economic and environmental consequences.

Linear and GLM bandits. For linear bandits with d -dimensional action/context space \mathcal{X} , the minimax expected cumulative regret is known to satisfy

$$\mathbb{E}[R_T] = O\left(d\sqrt{T}\log T\right),$$

where T is the time horizon (Abbasi-Yadkori et al., 2011; Dani et al., 2008). Generalized linear bandit models achieve similar rates up to logarithmic factors (Filippi et al., 2010; Bastani et al., 2020). These results provide a baseline: when yield or profit responds linearly to inputs, bandit algorithms can achieve fast learning rates. However, such linear approximations are rarely biologically realistic in crop response.

Nonparametric bandits. For more general nonparametric function classes, such as Lipschitz or RKHS (kernelized) reward functions, the minimax regret scales polynomially with the dimension d of the action/context space \mathcal{X} (Slivkins, 2014; Srinivas et al., 2012):

$$R_T = \tilde{O}\left(T^{\frac{d+1}{d+2}}\right).$$

While attractive for flexibility, these guarantees deteriorate with dimension and are impractical when only a small number of seasons or trials are available, which is a common reality in agriculture.

Expressive nonlinear models. Recent work has extended bandit algorithms to highly expressive nonlinear function classes such as neural networks. For example, NeuralUCB (Zhou et al., 2020) achieves $\tilde{O}(\sqrt{T})$ -type regret under effective-dimension assumptions. While theoretically appealing, such methods typically require substantially larger data and are less directly interpretable than the low-dimensional mechanistic models considered here. For completeness, we summarize Neural bandits in Appendix A and we also present representative regret rates for common stochastic bandit models in Table 3 of Appendix A.

While such results highlight the theoretical reach of modern methods, they are poorly suited to agricultural decision-making, where data are scarce, feedback is seasonal, and recommendations must be transparent to practitioners. Our focus, therefore, is on *mechanistic yield-response models* such as Mitscherlich or Michaelis-Menten. These models are parsimonious, interpretable, and biologically grounded, yet flexible enough to capture crop dose-response behavior. This specialization ensures theoretical guarantees translate into actionable insights, directly relevant for resource-efficient agricultural experimentation.

4.1 Model-based nonlinear bandits.

Building on these insights, we now focus on model-based nonlinear bandits, a regime particularly relevant in agronomy, where biological processes often admit interpretable, low-dimensional parameterizations. Since one of our motivations for this work is to provide decision-making algorithms in sample-limited scenarios, we focus on characterizing sample complexity more than providing regret guarantees in this section. Recent work by Dong et al. (2021) has shown a breakthrough for model-based nonlinear bandits: for any class of reward functions with bounded *sequential Rademacher complexity*, it is possible to find an ϵ -approximate local maximum with sample complexity polynomial in the complexity of the model class, *independent* of the action dimension. This suggests that, unlike classical methods whose sample complexity or regret is exponential in dimension, model-based methods can be vastly more efficient when the reward function admits a suitable low-complexity parameterization.

Sample Complexity and Local Regret for Model-based Nonlinear Bandits. In the model-based bandit setting, the functional form of the reward is known, but its parameters are not. The goal is therefore to efficiently identify an *approximate local maximum* of this parametric reward function, rather than a global maximum which is often computationally and statistically intractable. Here, *sample complexity* refers to the minimum number of experimental rounds required to guarantee, with high probability, that the recommended action achieves performance within a specified tolerance of a local optimum. This guarantee is formalized through the notion of *local regret*, which measures the suboptimality of the chosen action relative to the best locally optimal action. In practice, for well-behaved (e.g., unimodal) functions such as those common in agronomic yield response, a local maximum often coincides with the global maximum.

In particular, a point x is said to be an (ϵ_g, ϵ_h) -approximate local maximum if its gradient is at most ϵ_g (i.e., $\|\nabla f(x)\|_2 \leq \epsilon_g$) and the Hessian's largest eigenvalue is at most $-\epsilon_h$ (i.e., f is sufficiently concave around x). The sample complexity quantifies how quickly an algorithm can find such a point.

In this work, we focus on reward functions (e.g., quadratic-plateau, Michaelis-Menten, Mitscherlich, and logistic) that satisfy the regularity conditions (such as bounded gradients and Hessians) assumed in recent theoretical results (Dong et al., 2021). These properties ensure that our algorithms are well-behaved: the models are smooth enough for efficient learning, and the theoretical sample complexity guarantees apply. As a result, we can meaningfully compare algorithms in terms of how many rounds are needed to reach near-optimal fertilizer recommendations with high confidence. We first define the notion of sequential Rademacher complexity and then state Theorem 1.1 from Dong et al. (2021) in terms of sequential complexity, and then we tailor the result to our specific non-linear model classes.

Intuitively, sequential Rademacher complexity measures how hard it is for a learning algorithm to reliably make good decisions when facing an environment that can adapt to the algorithm's past actions. More formally, the sequential Rademacher complexity (Rakhlin et al., 2015b) is defined as follows.

Definition 1 (Sequential Rademacher Complexity). *Let \mathcal{F} be a class of real-valued functions defined on \mathcal{X} . The sequential Rademacher complexity of \mathcal{F} over T rounds is defined as*

$$\mathfrak{R}_T^{\text{seq}}(\mathcal{F}) = \sup_{x_1, \dots, x_T} \mathbb{E}_r \left[\sup_{f \in \mathcal{F}} \frac{1}{T} \sum_{t=1}^T r_t f(x_t(r_1, \dots, r_{t-1})) \right]$$

where the supremum is taken over all sequences of functions $x_t : \{-1, +1\}^{t-1} \rightarrow \mathcal{X}$, and r_1, \dots, r_T are independent Rademacher random variables, i.e., $\mathbb{P}(r_t = 1) = \mathbb{P}(r_t = -1) = \frac{1}{2}$ for each t .

Now, we present the result by Dong et al. (2021) that determines the sample complexity rate for `ViOlin`.

Theorem 1 (Sample Complexity for Model-based Nonlinear Bandits (Dong et al., 2021, Theorem 1.1)). *Suppose the sequential Rademacher complexity of the loss function class $\mathfrak{R}_T^{\text{seq}}(\mathcal{F})$ induced by the reward function class $\{f(\theta, \cdot) : \theta \in \Theta\}$ is bounded by $\sqrt{R(\Theta)T \text{polylog}(T)}$ for some complexity parameter $R(\Theta)$. Then, there exists an algorithm (`ViOlin`) (Algorithm 4) that finds an δ -approximate local maximum with $\tilde{O}(R(\Theta)\delta^{-8})$ samples, independent of the dimension of the action space.*

Note that most of the agronomic models assume bounded non-linear functions, such as ones that exhibit plateauing behavior after a certain threshold. Therefore, it is of interest to quantify the sequential Rademacher complexity for these class of functions. For parametric classes with bounded functions and inputs (such as those considered in this paper), we prove a theoretical bound on the sequential Rademacher complexity of bounded functions and show that indeed for this class of functions the sequential Rademacher complexity has an upper bound as required in Theorem 1.

Theorem 2 (Sequential Rademacher Complexity for Bounded Functions). *Let $\mathcal{F} \subseteq [-B_{\mathcal{F}}, B_{\mathcal{F}}]^{\mathcal{D}}$ be a class of functions uniformly bounded by $B_{\mathcal{F}} > 0$. Then the sequential Rademacher complexity of \mathcal{F} satisfies*

$$\mathfrak{R}_T^{\text{seq}}(\mathcal{F}) := \sup_x \mathbb{E}_r \left[\sup_{f \in \mathcal{F}} \frac{1}{T} \sum_{t=1}^T r_t f(x_t(r_1, \dots, r_{t-1})) \right] \leq C \cdot B_{\mathcal{F}} \cdot \sqrt{\frac{\log T}{T}},$$

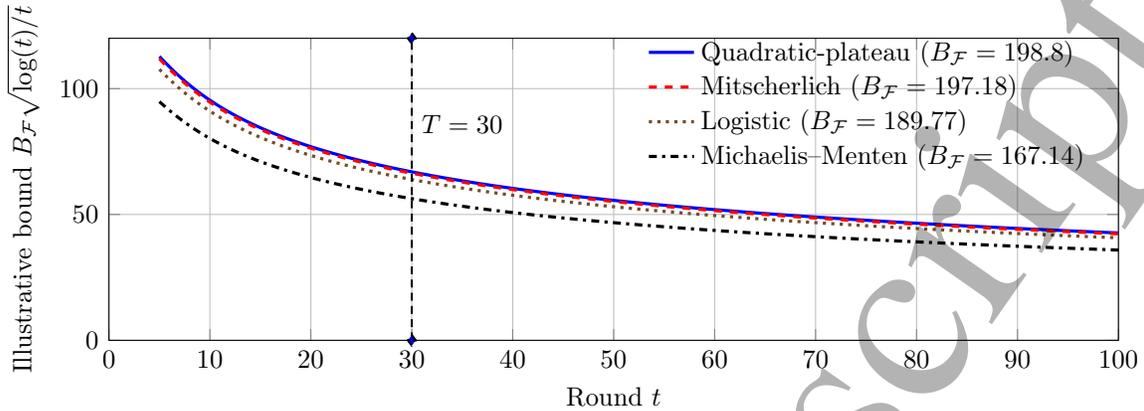


Figure 1: **Illustration of model-dependent upper bounds on sequential Rademacher complexity.** We plot $B_{\mathcal{F}} \sqrt{\log(t)}/t$ (up to a universal constant common to all models) for the four yield-response families used in the simulations, where $B_{\mathcal{F}} = \sup_{x \in [0, 250]} |f(x; \theta_{\text{true}})|$ is computed from the parameter settings in Section 5. The bound decreases with the number of rounds t ; differences across model families appear through the constant $B_{\mathcal{F}}$. The dashed vertical line marks the horizon $T = 30$ used in the well-specified experiments.

for some universal constant $C > 0$.

We defer the proof of Theorem 2 to the Appendix (Section B) for brevity and discuss its implications from an application viewpoint. Our theoretical result complement Dong et al. (2021)’s general framework by concretely instantiating the sample complexity for these practically important nonlinear classes, including quadratic plateau (threshold), Michaelis-Menten, Mitscherlich, and logistic models, the sequential Rademacher complexity $\mathfrak{R}_T^{\text{seq}}(\mathcal{F})$ can be bounded as $\tilde{O}(1/\sqrt{T})$. Figure 1 visualizes the bound in Theorem 2 for the four nonlinear yield-response families in Table 2. In this result, model dependence enters through a uniform magnitude bound $B_{\mathcal{F}}$ satisfying $\sup_{x \in \mathcal{X}} |f(x)| \leq B_{\mathcal{F}}$ for all $f \in \mathcal{F}$. We set $B_{\mathcal{F}} = \sup_{x \in [0, 250]} |f(x)|$ using the parameter settings from Section 5 and plot the resulting (conservative) upper bounds versus t ; the corresponding $B_{\mathcal{F}}$ values are reported in Appendix B.1. This figure is included purely for intuition: it illustrates that the (conservative) complexity bound decreases as the number of rounds increases, and it should not be interpreted as a model-selection criterion. Therefore, the sample complexity for identifying an ϵ -optimal arm (in the sense of a local maximum) in these bandit settings scales as $\tilde{O}(1/\epsilon^8)$, with constants depending on the parameter and input bounds. This bound, as established in Dong et al. (2021), is general and not specialized to the structure of agronomic models in terms of its polynomial dependence on ϵ . Deriving sharper rates for the structured nonlinear models considered here is an important direction for future work.

Corollary 1 (Sample Complexity for online learning of Bounded Non-linear Reward classes). *Using the bound in Theorem 2 in Theorem 1, an ϵ -optimal solution (arm) can be identified with sample complexity (number of time steps), $\tilde{O}(1/\delta^8)$ for the *Violin* algorithm of (Dong et al., 2021), independent of the action (arm) space dimension.*

Interpretation. This shows that for commonly used nonlinear agronomic yield response models, bandit algorithms can efficiently identify near-optimal fertilizer rates with a number of field trials that scales polynomially with the desired accuracy and avoids the curse of dimensionality in the action space. Our results instantiate Dong et al. (2021)’s general theory for these specific models,

providing practical sample complexity bounds for adaptive experimentation in agriculture. In Theorem 1, the parameter $\delta > 0$ represents the maximum allowable gap between the algorithm’s selected action (e.g., a nitrogen rate recommendation) and the optimal (locally maximizing) action in terms of expected reward. That is, with high probability, the bandit algorithm identifies an action x such that

$$f^*(x^*) - f^*(x) \leq \delta,$$

where f^* is the true (unknown) reward or profit function, x^* is a (local) maximizer of f^* , and x is the action recommended by the algorithm.

The sample complexity required to achieve an ϵ -optimal recommendation scales polynomially with $1/\delta$ (specifically, as $O(1/\delta^8)$ in Theorem 1), so tighter tolerances (smaller δ) require proportionally more field trials. In practice, δ should be set according to what constitutes an economically meaningful margin for decision-making.

Dong et al. (2021) provide broad sample complexity and regret guarantees for nonlinear bandit problems, including highly expressive classes such as 2-layer neural networks. By contrast, our agricultural decision models are low-dimensional and mechanistic (e.g., quadratic plateau), which makes them both interpretable and practically relevant. In this structured setting, algorithms such as UCB and ϵ -greedy are not only straightforward to implement but also expected to admit stronger theoretical guarantees than the general results in Dong et al. (2021). We adopt their bounds as a reference baseline, while highlighting the opportunity for future work to establish sharper, model-specific results. This positions our empirical study on firm theoretical footing while opening avenues for more refined analysis.

Remark 3. *The proof techniques used here such as bounding regret via sequential Rademacher complexity and the resulting sample-complexity corollary for `Violin` are direct instantiations of general tools from the bandit literature. What is distinctive in this result is their adaptation to mechanistic nonlinear models (e.g., Mitscherlich, Michaelis-Menten), which integrates domain knowledge with statistical guarantees. This integration ensures that theoretical results are not only mathematically sound but also directly interpretable and actionable in agronomic applications.*

Under Model Misspecification In many applications, including agricultural yield optimization, the true reward function may not be perfectly captured by any member of a chosen parametric model class (such as quadratic plateau or Michaelis-Menten). This is known as *model misspecification*, and can have important implications for the performance and guarantees of bandit algorithms. A concrete example in agronomy is when the true yield response exhibits smooth saturation (well captured by a Mitscherlich curve), but the fitted model class is a quadratic-plateau response. In that case, the quadratic-plateau model may fit the central range reasonably well but can misrepresent the curvature near the shoulder/plateau region, leading to biased estimates of the profit-maximizing nitrogen rate.

Most bandit algorithms rely on the assumption of a *well-specified model* (or “realizability”), that is, the existence of a function $f^* \in \mathcal{F}$ such that the mean reward is $\mu(x) = f^*(x)$ for all actions x (Chu et al., 2011; Abbasi-Yadkori et al., 2011; Agarwal et al., 2012; Foster et al., 2018). However, in practice, exact realizability rarely holds, and it is important to consider the effect of model misspecification.

Recent work (see, e.g., (Greenewald et al., 2021; Lattimore et al., 2020; Ghosh et al., 2017)) has studied bandits under various notions of misspecification. A common formulation is the *uniform*

ε -misspecification setting, where

$$\inf_{f \in \mathcal{F}} \sup_{x \in \mathcal{X}} |\mu(x) - f(x)| \leq \varepsilon_{\mathcal{F}}, \quad (8)$$

for some $\varepsilon_{\mathcal{F}} > 0$.

For linear bandits, Lattimore et al. (2020) show that the cumulative regret must satisfy

$$R_T \gtrsim d\sqrt{T} + \varepsilon_{\mathcal{F}}\sqrt{dT},$$

where d is the model dimension and T is the time horizon. The first term corresponds to the minimax regret under well-specified models, while the second is an unavoidable “price of misspecification”. In our setting, the models (quadratic plateau, logistic, etc.) may be misspecified relative to the true reward. As such, the regret of our bandit algorithms can be interpreted as

$$R_T = O(\text{model-based regret}) + O(T \varepsilon_{\mathcal{F}}),$$

where $\varepsilon_{\mathcal{F}}$ is the uniform misspecification level in (8). This highlights the need for robust algorithms and motivates future work on model selection or hybrid model-based/nonparametric approaches.

5 Simulation study: Emulating agricultural decision-making

5.1 Experimental setup and evaluation metrics

Building on the profit objective in Section 2.1, we evaluate the proposed bandit algorithms in controlled simulations that mimic fertilizer-rate decisions under limited data. In each experiment, the mean yield response is specified by a chosen model family $f(x; \theta)$ (Mitscherlich, quadratic-plateau, Michaelis–Menten, or logistic), actions are restricted to a discrete grid \mathcal{X} of nitrogen rates, and economic conditions are varied through the price parameters (p_y, p_x) . At any round, if the parameters θ were known, the profit-maximizing rate would be

$$x^* \in \arg \max_{x \in \mathcal{X}} \Pi(x) = \arg \max_{x \in \mathcal{X}} \{p_y f(x; \theta) - p_x x\}.$$

In practice, θ is unknown and must be learned online from noisy outcomes, so the algorithms repeatedly fit a model and select fertilizer rates based on plug-in profit estimates and exploration bonuses. For the nonlinear response families considered here, the continuous maximizer admits closed-form expressions (Table 2). These formulas clarify how the economically optimal decision depends jointly on agronomic parameters θ and prices (p_y, p_x) , motivating regret and profit-based comparisons as data accumulate over rounds.

We now turn to the problem of sequentially optimizing fertilizer decisions using model-based bandit algorithms that account for uncertainty and limited data. At each round, the algorithm fits a nonlinear yield response model to the observed data and selects the fertilizer rate that maximizes the estimated economic profit, optionally incorporating an uncertainty-based exploration bonus. We investigate three nonlinear model-based strategies below. To avoid confusion with the generic reward (yield) formulation in Section 3, Algorithms 5–7 in this section are profit-based variants of Algorithms 2–4 (and the baselines), obtained by replacing the reward with the profit objective $\Pi(x) = p_y Y(x) - p_x x$. Algorithms 2–4 provide the core reward-based pseudocode templates, while Algorithms 5–7 give the corresponding simulation-ready, implementation-level versions used in Section 5.

Table 2: **Closed-form fertilizer rate that maximizes profit under four yield–nitrogen response models.** We maximize per-round profit $\Pi(x) = p_y Y(x) - p_x x$, where x is the nitrogen rate (lb N/ac), $Y(x)$ is yield, p_y is grain price (\$/bu), and p_x is nitrogen price (\$/lb N). Model parameters are defined in Section 5. In the simulations, actions are chosen from a discrete grid \mathcal{X} (e.g., $\{0, 50, \dots, 250\}$ lb N/ac), so the implemented decision is the nearest grid value to the continuous maximizer reported here.

Model	Yield Function $Y(x)$	Profit-maximizing x^*
Mitscherlich	$A(1 - e^{-bx})$	$-\frac{1}{b} \ln\left(\frac{p_x}{p_y A b}\right)$
Quadratic (threshold)	$a + bx + cx^2, x \leq x_0$ $a + bx_0 + cx_0^2, x > x_0$	$\min\left\{x_0, \max\left\{0, \frac{1}{2c}\left(\frac{p_x}{p_y} - b\right)\right\}\right\}$
Michaelis-Menten	$\frac{ax}{b+x}$	$\sqrt{\frac{abp_y}{p_x}} - b$
Logistic	$\frac{A}{1 + e^{-B(x-C)}}$	$u^* = \frac{\gamma - 2 - \sqrt{(\gamma - 2)^2 - 4}}{2},$ $\gamma = \frac{Bp_y A}{p_x}, \quad x^* = C - \frac{1}{B} \ln u^*$

1. *Model-based ϵ -greedy algorithm (Algorithm 5)*: This method selects a random fertilizer rate with probability $\epsilon_t = t^{-a}$, with a chosen from $\{0.5, 1, 1.5\}$, for exploration. With the remaining $(1 - \epsilon_t)$ probability, it chooses the rate closest to the current profit-maximizing estimate x^* , computed using the fitted model and closed-form expressions where available.
2. *Model-based UCB algorithm (Algorithm 6)*: This approach augments the estimated profit at each arm with a model-based uncertainty term, derived from the parameter covariance of the fitted nonlinear model, and selects the rate with the highest upper confidence bound. Model based nonlinear-UCB uses the delta-method uncertainty proxy (6) with warm-start n_0 .
3. *Violin algorithm (Algorithm 4)*: A curvature-aware strategy that exploits local second-order structure in the profit function to accelerate convergence, particularly under smooth nonlinearity.

Although these methods differ in their exploration strategies, they all share the same structure: fitting a nonlinear model to guide adaptive decision-making under uncertainty. Accurate initialization of nonlinear model parameters is crucial to the success of these algorithms, especially in small-sample settings. To benchmark the performance of nonlinear model-based strategies, we compare them against two widely used baselines: LinUCB (Chu et al., 2011) and k -Nearest Neighbor-UCB (kNN-UCB) (Reeve et al., 2018).

4. *LinUCB algorithm (Algorithm 8)*: A linear bandit method that models yield as a linear function of fertilizer rate and selects actions using upper confidence bounds derived from online regression.
5. *kNN-UCB algorithm (Algorithm 9)*: A nonparametric method that estimates yield using the k nearest neighbors of each candidate arm and quantifies uncertainty via local sample variance.

Further descriptions of these algorithms can be found in the Appendix. These baselines span the spectrum from parametric simplicity to nonparametric flexibility and help highlight the advantages

Algorithm 5 Model-Based ϵ -Greedy for Economic Profit Maximization

Require: Fertilizer levels \mathcal{X} , time horizon T , prices p_y, p_x , exploration rate ϵ_t

```

1: Initialize dataset  $\mathcal{D} \leftarrow \emptyset$ 
2: for  $t = 1$  to  $T$  do
3:   if  $u \sim \text{Uniform}(0, 1) < \epsilon_t$  or  $|\mathcal{D}| < n_0$  then
4:     Select fertilizer level  $x_t \sim \text{Uniform}(\mathcal{X})$  // Explore
5:   else
6:     Fit model  $f(x; \hat{\theta})$  to data in  $\mathcal{D}$ 
7:     Compute  $x^* = \arg \max_{x \in \mathcal{X}} [p_y f(x; \hat{\theta}) - p_x x]$ 
8:     Select  $x_t \in \mathcal{X}$  closest to  $x^*$ 
9:   end if
10:  Apply  $x_t$ , observe yield  $Y_t$ 
11:  Compute profit:  $\Pi_t = p_y \cdot Y_t - p_x \cdot x_t$ 
12:  Add  $(x_t, Y_t)$  to dataset  $\mathcal{D}$ 
13: end for

```

Algorithm 6 Model-Based UCB for Economic Profit Maximization

Require: Fertilizer levels \mathcal{X} , horizon T , prices p_y, p_x , UCB constant α , warm-start length n_0

```

1: Initialize dataset  $\mathcal{D} \leftarrow \emptyset$ 
2: for  $t = 1$  to  $T$  do
3:   if  $|\mathcal{D}| < n_0$  then
4:     Select  $x_t \sim \text{Uniform}(\mathcal{X})$  // warm-start exploration
5:   else
6:     Fit reward model  $f(x; \hat{\theta})$  to data  $\mathcal{D}$ 
7:     Estimate parameter covariance  $\widehat{\text{Cov}}(\hat{\theta})$  (e.g., from nonlinear least squares)
8:     for each  $x \in \mathcal{X}$  do
9:       Predict profit:  $\hat{\Pi}(x) = p_y \cdot f(x; \hat{\theta}) - p_x \cdot x$ 
10:      Compute uncertainty proxy (delta-method, cf. (6)):
11:
12:
13:
14:
15:
16:
17:
18:
19:
20:
21:
22:
23:
24:
25:
26:
27:
28:
29:
30:
31:
32:
33:
34:
35:
36:
37:
38:
39:
40:
41:
42:
43:
44:
45:
46:
47:
48:
49:
50:
51:
52:
53:
54:
55:
56:
57:
58:
59:
60:

```

$$\text{Unc}(x) = p_y \cdot \sqrt{\nabla_{\theta} f(x; \hat{\theta})^{\top} \widehat{\text{Cov}}(\hat{\theta}) \nabla_{\theta} f(x; \hat{\theta})}$$

```

11:      Compute UCB score:  $UCB(x) = \hat{\Pi}(x) + \alpha \cdot \text{Unc}(x)$ 
12:    end for
13:    Select  $x_t = \arg \max_{x \in \mathcal{X}} UCB(x)$ 
14:  end if
15:  Apply  $x_t$ , observe yield  $Y_t$ 
16:  Compute profit:  $\Pi_t = p_y \cdot Y_t - p_x \cdot x_t$ 
17:  Update  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x_t, Y_t)\}$ 
18: end for

```

of domain-informed nonlinear models in sample-constrained settings.

In the next section, we present simulation studies under both well-specified and misspecified conditions to compare the performance of these five bandit algorithms in terms of profit and cumulative regret.

Algorithm 7 Vi01in for Economic Profit Maximization

Require: Fertilizer levels \mathcal{X} , horizon T , prices p_y, p_x , curvature weights $\kappa_1, \kappa_2 \geq 0$, warm-start length n_0 , learner/estimator \mathcal{O}

1: Initialize dataset $\mathcal{D} \leftarrow \emptyset$

2: **for** $t = 1$ to T **do**

3: **if** $|\mathcal{D}| < n_0$ **then**

4: Select $x_t \sim \text{Uniform}(\mathcal{X})$

// warm-start exploration

5: **else**

6: Fit/update $\hat{\theta}$ using \mathcal{D} via \mathcal{O}

7: **for** each $x \in \mathcal{X}$ **do**

8: Predict profit: $\hat{\Pi}(x) = p_y f(x; \hat{\theta}) - p_x x$

9: Compute profit-gradient and curvature terms:

$$\hat{g}_{\Pi}(x) = \partial_x \hat{\Pi}(x) = p_y \partial_x f(x; \hat{\theta}) - p_x, \quad \hat{H}_{\Pi}(x) = \partial_{xx}^2 \hat{\Pi}(x) = p_y \partial_{xx}^2 f(x; \hat{\theta})$$

10: Compute Vi01in score:

$$S_t(x) = \hat{\Pi}(x) + \kappa_1 |\hat{g}_{\Pi}(x)| + \kappa_2 |\hat{H}_{\Pi}(x)|$$

11: **end for**

12: Select $x_t \in \arg \max_{x \in \mathcal{X}} S_t(x)$

13: **end if**

14: Apply x_t , observe yield Y_t

15: Compute profit: $\Pi_t = p_y Y_t - p_x x_t$

16: Update $\mathcal{D} \leftarrow \mathcal{D} \cup \{(x_t, Y_t)\}$

17: **end for**

5.2 Well-specified setting

Simulation setup. In the well-specified setting, data were generated according to a quadratic plateau model of the form:

$$Y(x) = \begin{cases} a + bx + cx^2, & \text{if } x \leq x_0 \\ a + bx_0 + cx_0^2, & \text{if } x > x_0 \end{cases},$$

where $Y(x)$ denotes crop yield (bu/ac) at fertilizer rate x (lb N/ac), and $a = 80$, $b = 1.2$, $c = -0.003$, $x_0 = 180$ were chosen to reflect realistic agronomic response curves. The profit at each round is computed as $\Pi(x) = p_y Y(x) - p_x x$, with grain price $p_y = \$5.00$ per bushel and fertilizer price $p_x \in \{0.3, 0.5, 0.7\}$ \$/lb N to mimic realistic prices per unit nitrogen fertilizer (such as Urea) in the Midwest US. Fertilizer rates are restricted to the discrete grid $\mathcal{X} = \{0, 50, 100, \dots, 250\}$ lb N/ac, and yields are corrupted by i.i.d. Gaussian noise with standard deviation $\sigma = 0.5$.

We evaluated five algorithms: model-based ϵ -greedy, nonlinear UCB, Vi01in (curvature-matched nonlinear bandit), LinUCB (linear model-based UCB), and kNN-UCB (nonparametric UCB using $k = 3$ nearest neighbors). Hyperparameters for each of the algorithms are chosen based on theoretical guidelines and grid searches over multiple runs. We choose $\epsilon_t = t^{-1.5}$ for ϵ -greedy in this setting, $\alpha = 1$ in UCB, linUCB and kNN-UCB for uniformity and fairness in comparison, $\alpha_1 = 2.0, \alpha_2 = 640$ for the Vi01in algorithm based on guidelines in Dong et al. (2021). For each algorithm and parameter configuration, we performed 10 independent simulation replicates

Algorithm 8 LinUCB for Economic Profit Maximization

Require: Fertilizer levels \mathcal{X} , time horizon T , prices p_y, p_x , UCB parameter α

```

1: Initialize dataset  $\mathcal{D} \leftarrow \emptyset$ 
2: Define feature mapping  $\varphi(x) = [1, x]^\top$ 
3: for  $t = 1$  to  $T$  do
4:   if not enough data then
5:     Choose  $x_t$  uniformly at random from  $\mathcal{X}$ 
6:   else
7:     Fit linear model to data:  $Y_i \approx \varphi(x_i)^\top \hat{\beta}$ 
8:     For each  $x \in \mathcal{X}$ :
9:       • Predict yield:  $\hat{Y}(x) = \varphi(x)^\top \hat{\beta}$ 
10:      • Compute profit:  $\hat{\Pi}(x) = p_y \cdot \hat{Y}(x) - p_x \cdot x$ 
11:      • Estimate variance:  $s^2(x) = \varphi(x)^\top (\mathbf{V}^{-1}) \varphi(x)$ , where  $\mathbf{V}$  is the feature covariance
12:      matrix
13:      • Compute UCB score:  $UCB(x) = \hat{\Pi}(x) + \alpha \cdot s(x)$ 
14:     Select  $x_t = \arg \max_{x \in \mathcal{X}} UCB(x)$ 
15:   end if
16:   Apply  $x_t$ , observe  $Y_t$ 
17:   Compute profit:  $\Pi_t = p_y \cdot Y_t - p_x \cdot x_t$ 
18:   Append  $(x_t, Y_t)$  to  $\mathcal{D}$ 
19: end for

```

Algorithm 9 kNN-UCB for Economic Profit Maximization

Require: Fertilizer levels \mathcal{X} , time horizon T , prices p_y, p_x , UCB parameter α , no. of neighbors k

```

1: Initialize dataset  $\mathcal{D} \leftarrow \emptyset$ 
2: for  $t = 1$  to  $T$  do
3:   if not enough data then
4:     Choose  $x_t$  uniformly at random from  $\mathcal{X}$ 
5:   else
6:     for each  $x \in \mathcal{X}$  do
7:       Find  $k$  nearest previously tried fertilizer rates to  $x$  in  $\mathcal{D}$ 
8:       Compute average yield:  $\hat{Y}_k(x) = \frac{1}{k} \sum_{j=1}^k Y_j$ 
9:       Compute sample standard deviation:  $s_k(x)$  of yields among neighbors
10:      Compute profit:  $\hat{\Pi}_k(x) = p_y \cdot \hat{Y}_k(x) - p_x \cdot x$ 
11:      Compute UCB score:  $UCB(x) = \hat{\Pi}_k(x) + \alpha \cdot \frac{s_k(x)}{\sqrt{k}}$ 
12:     end for
13:     Select  $x_t = \arg \max_{x \in \mathcal{X}} UCB(x)$ 
14:   end if
15:   Apply  $x_t$ , observe  $Y_t$ 
16:   Compute profit:  $\Pi_t = p_y \cdot Y_t - p_x \cdot x_t$ 
17:   Append  $(x_t, Y_t)$  to  $\mathcal{D}$ 
18: end for

```

of $T = 30$ rounds. Each algorithm fits its specified model to the observed data and selects actions to maximize estimated profit. The regret at each round is defined as the difference between the average profit for the optimal arm (computed using knowledge of the true reward function) and

the average profit realized by the arm selected by the algorithm, that is,

$$\text{Regret}_t = \mathbb{E} [\Pi(x^*; \theta^*) - \Pi(x_t; \theta^*)],$$

where x^* denotes the arm maximizing the true profit function, θ^* the true model parameters and x_t is the arm selected at round t by the bandit algorithm employed to make decisions. This definition of regret translates directly into wasted profit and fertilizer input, making it a practical metric for resource-efficient agricultural decision-making. For each run, we record cumulative regret, average profit, and select fertilizer rates. In order to quantify the uncertainty, we also plot the distributions of cumulative regret and average profit over 10 replications at round 10, 20, and 30, respectively. Note, profit, regret and all other monetary quantities are reported in \$/ac (yield in bu/ac and nitrogen in lb/ac) unless stated otherwise.

Results and Interpretation. Figure 2 presents the average cumulative regret and the average profit per round for each algorithm. Figure 3 presents boxplots of the cumulative regret and average profit for each algorithm, evaluated at rounds 10, 20, and 30, respectively, across 10 independent replications. Several trends are immediately apparent:

- **Superiority of nonlinear model-based methods:** Both nonparametric (kNN-UCB) and linear (LinUCB) methods incur substantially higher regret than the nonlinear model-based approaches. In particular, kNN-UCB and LinUCB accumulate regret rapidly and fail to converge to near-optimal recommendations within the small sample budget. LinUCB, although it does not experience a sharp initial dip in profit due to exploration, tends to settle on a suboptimal arm and consistently incurs losses relative to the optimal strategy.

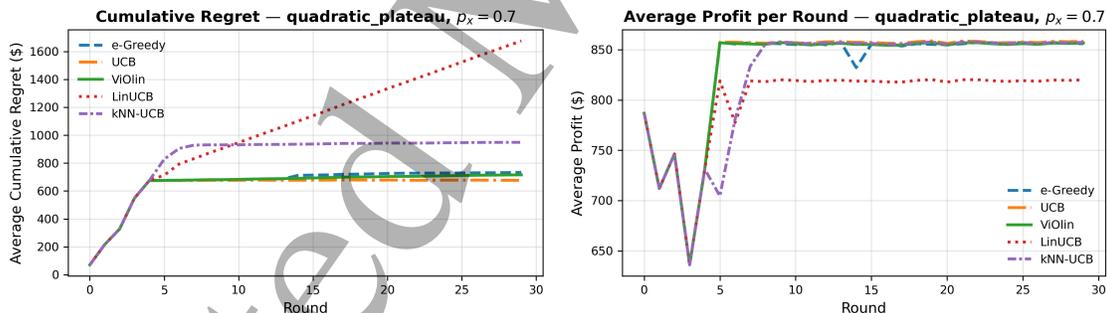


Figure 2: **Well-specified quadratic-plateau profit experiment.** Actions lie in $\mathcal{X} = \{0, 50, \dots, 250\}$ lb N/ac with $p_y = \$5/\text{bu}$, $p_x = \$0.7/\text{lb N}$, $T = 30$, and $\sigma = 0.5$. **Left:** cumulative profit regret (\$/ac). **Right:** average profit (\$/ac). Curves show means over 10 replicates for ϵ -greedy, nonlinear-UCB, ViOlin, LinUCB, and kNN-UCB; ViOlin and nonlinear-UCB reduce regret fastest.

- **Exploration versus exploitation trade-off:** All algorithms that actively explore—notably ϵ -greedy and UCB—show a pronounced dip in profit at early rounds, reflecting the cost of exploratory actions. In contrast, the ViOlin algorithm, which is designed to be more exploitative (greedily maximizing predicted profit with curvature matching), avoids this initial dip and achieves high profits almost immediately. However, both ϵ -greedy and UCB eventually recover and converge to policies with low cumulative regret, validating the effectiveness of their exploration in learning the optimal fertilizer rate.

- Uncertainty Quantification:** Figure 3 highlights important differences between model-based approaches (e.g., Violin, model based UCB) against linear and nonparametric benchmarks such as lin-UCB and kNN-UCB, particularly in the small-sample regime relevant to agricultural field trials. While kNN-UCB exhibits relatively low variability across replicates, suggesting stable short-term performance, this stability arises primarily from smoothing rather than from capturing the underlying input-response mechanism. As a result, kNN-UCB can suffer from persistent bias, leading to consistently suboptimal mean performance. This is evident in the profit distributions, i.e., the mean profit for kNN-UCB consistently lies below the lower quartile of the nonlinear model-based approaches. In other words, the algorithm produces profits that are reliably clustered, but around a lower and biased center. This suggests that while its variability is small, its systematic bias leads to consistently suboptimal performance.

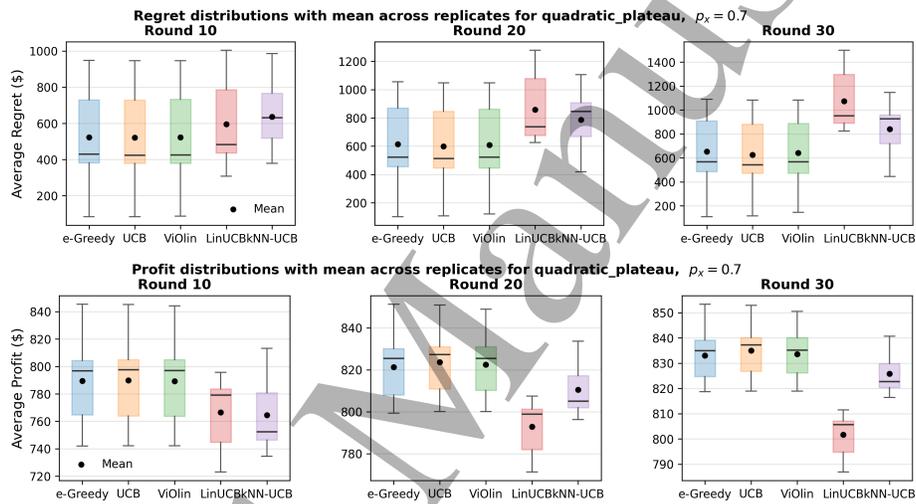


Figure 3: **Between-run variability for the well-specified quadratic-plateau experiment.** Boxplots summarize results across 10 independent simulation replicates at rounds $t \in \{10, 20, 30\}$. **Top row:** cumulative regret (in \$/ac). **Bottom row:** average per-round profit (in \$/ac). Each box shows the interquartile range (25th–75th percentiles) with the median; whiskers indicate the spread; the dot marks the mean.

In contrast, nonlinear model-based algorithms exhibit higher run-to-run variability due to parameter estimation uncertainty, particularly in early rounds when data are scarce. However, these algorithms leverage structural assumptions that align with biological processes, enabling faster convergence to the true optimum as more data accrue.

Similar trends were observed across other values of the fertilizer price p_x and for all nonlinear response models considered. Taken together, these simulation studies demonstrate a consistent advantage of nonlinear model-based bandit algorithms over both nonparametric and linear parametric alternatives in small-sample regimes. Based on these findings, we subsequently focus our comparisons on the family of nonlinear model-based algorithms, investigating their relative strengths under a range of simulation scenarios to clarify which approaches are most advantageous in which settings.

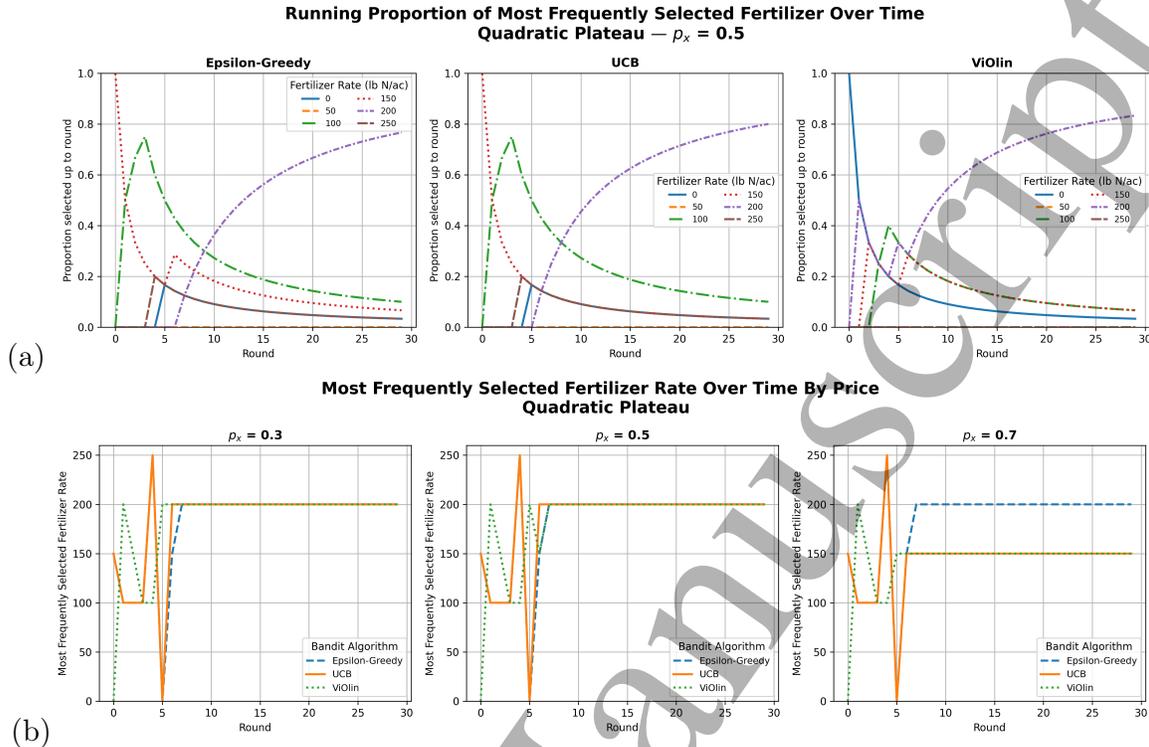


Figure 4: **How the algorithms' fertilizer-rate choices evolve over time (well-specified quadratic plateau).** Actions are chosen from $\mathcal{X} = \{0, 50, \dots, 250\}$ lb N/ac. (a) For $p_x = \$0.5/\text{lb N}$, the running proportion of times each nitrogen rate has been selected up to round t is shown for ϵ -greedy, nonlinear-UCB, and Violin (averaged over 10 replicates). (b) The most frequently selected nitrogen rate at each round is shown for fertilizer prices $p_x \in \{0.3, 0.5, 0.7\}$ $\$/\text{lb N}$, illustrating how higher fertilizer cost shifts the learned decision toward lower nitrogen rates.

In this next part of the simulation study, we focus on illustrating the implications of using the three non-linear model-based bandit algorithms for learning fertilizer rates sequentially with the goal of profit maximization. Again in a small-sample regime, we run each of our algorithms for $T = 30$ rounds and replicate each run 10 times. Since our algorithms choose arms over time, it is important to visualize how the arm choices evolve over time for each of the algorithms. In Figure 4(a), we plot the running proportion of fertilizer rate ranging from $\{0, 50, \dots, 250\}$ lb N/ac, selected at each round by each of the algorithm over the 10 replicates for the quadratic plateau model for $p_x = \$0.5/\text{lb N}$ to mimic realistic prices per unit nitrogen fertilizer (Urea) in the Midwest US. Note that in the beginning, the UCB and ϵ -greedy ($\epsilon_t = t^{-1.5}$) explore other arms such as $x = 100, 150$ lb N/ac, but just after a few rounds of exploration, learn the better arm to be $x = 200$ lb N/ac. Violin is more greedy from the beginning itself as can be seen by the proportions of other arms (other than purple) selected in the initial rounds. In Figure 4(b), in order to assess the affect of increasing fertilizer prices on decision-making for profit maximization, we also plotted the average selected fertilizer rate over time for the three algorithms. We note that the optimum arm (fertilizer rate) choice decreases as p_x (price per unit) increases from $\$0.3$ to $\$0.7/\text{lb N}$.

Additionally, since our decisions depend significantly on how well our parameters for the non-linear model are estimated over time, in Figure 5, we also plot the parameter trajectories over time for the quadratic plateau model with $p_x = \$0.5/\text{lb N}$. Note that all the three algorithms, as data

accumulates, the parameter estimates stabilize close to the true parameter values that were used to generate the data. In terms of interpretability, one can, for example, look at the parameter estimates for x_0 (bottom right), which denotes the threshold value at which the quadratic model plateaus, thus reflecting the saturating point over rounds (or seasons), beyond which additional fertilizer amount provide little benefit. In our simulations, we also examined the sensitivity of parameter estimation to initialization. Because nonlinear least-squares fitting is iterative, the choice of starting values can influence convergence, particularly in small-sample regimes. To reflect realistic practice, we initialized parameters at values close to but not equal to the true generating parameters (see Section E in the Appendix for explicit values used in each model). We observed that biologically plausible initializations led to faster and more stable convergence, while poor starting values could slow parameter recovery. In practice, we recommend using domain knowledge (e.g., agronomic ranges for maximum yield, response rates, or plateau points) to set initial values when fitting nonlinear models. We conducted similar experiments for the other non-linear models and those results are presented in the Appendix.

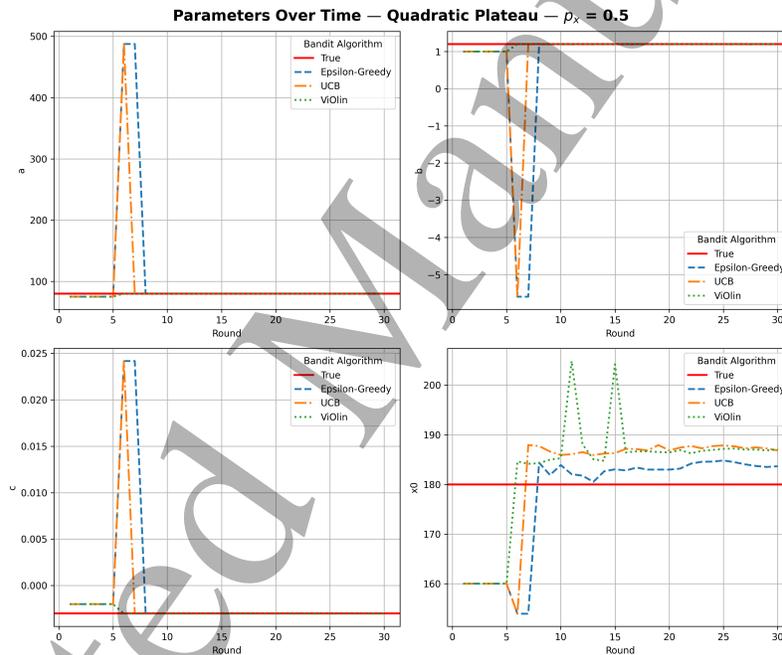


Figure 5: **Model-parameter learning over time (well-specified quadratic-plateau).** For $p_x = \$0.5/\text{lb N}$ and $T = 30$, we plot estimated (a, b, c, x_0) for ϵ -greedy, nonlinear-UCB, and ViOlin; The horizontal red line denotes the true value. Parameter estimates stabilize over time, indicating that the response curve can be learned from sequential data within the horizon considered. (One representative replicate is shown; similar behavior occurs across runs.)

5.3 Misspecified setting

Simulation Setup. To mimic a realistic situation in which the decision-maker fits an *approximate* but shape-compatible model, we generated data from the Mitscherlich response

$$Y_{\text{true}}(x) = A \left(1 - e^{-b(x-d)} \right), \quad A = 120, \quad b = 0.015, \quad d = 80,$$

yet *fitted* a quadratic-plateau curve

$$Y_{\text{fit}}(x) = a + bx + cx^2 \text{ for } x \leq x_0,$$

truncated at $x_0 = 180$ and takes the same value $a + bx_0 + cx_0^2$ for $x > x_0$. All other ingredients are identical to the well-specified experiment except that we increase the time horizon to $T = 100$, specifically, price grid $p_x \in \{0.3, 0.5, 0.7\}$ $\$/\text{lb}^{-1}$ N, yield price $p_y = \$5.00 \text{ bu}^{-1}$, action set $\mathcal{X} = \{0, 50, 100, \dots, 250\}$ lb N /ac , noise $\sigma = 0.5$, and the 5 algorithms: ϵ -greedy, nonlinear-UCB, ViOlin, LinUCB, kNN-UCB. 10 independent replicates were run for each algorithm and price.

Figure 6 plots the average cumulative regret when the fitted quadratic-plateau and Michaelis Menten model is misspecified with respect to the true Mitscherlich process. Relative to the well-specified case, the model-based regret curves shift upwards, reflecting the price of fitting an approximate response surface. Yet their slopes remain moderate, indicating that the fitted quadratic-plateau still guides the search toward profitable regions quickly. Note, that similar to the well-specified setting, the model-based non-linear bandits perform better than the linear and non-parametric baselines.

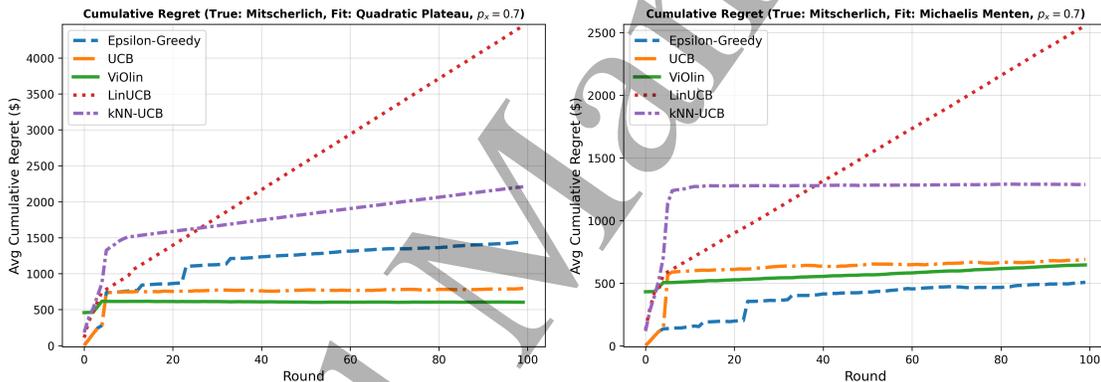


Figure 6: **Profit regret under model misspecification.** Data are generated from a Mitscherlich (truth) yield curve, but the learner fits a different parametric family. We use $p_y = \$5/\text{bu}$, $p_x = \$0.7/\text{lb N}$, $\mathcal{X} = \{0, 50, \dots, 250\}$ lb N /ac , $\sigma = 0.5$, and $T = 100$. **Left:** quadratic-plateau fit; **Right:** Michaelis–Menten fit. Curves show mean cumulative profit regret (in $\$/\text{ac}$) over 10 replicates for ϵ -greedy, nonlinear-UCB, ViOlin, LinUCB, and kNN-UCB. Under misspecification, regret increases for all methods, but nonlinear model-based policies remain competitive.

Although the quadratic-plateau form cannot capture the exponential approach to an asymptote exactly, its concave-then-flat shape matches the broad geometry of the Mitscherlich curve. As a consequence, all three nonlinear model-based algorithms, ϵ -greedy, UCB, and ViOlin continue to accumulate substantially *less* regret than the linear (LinUCB) or non-parametric (kNN-UCB) baselines. These results highlight the resilience of shape-compatible models under misspecification and point to rich opportunities for deeper theoretical and empirical study. Although ViOlin’s greedier, curvature-penalized score again avoids the large initial dip seen in UCB and ϵ -greedy, note that UCB and ϵ -greedy seem to perform comparably, especially in the Michaelis Menten fit, perhaps suggesting that exploration might help in misspecified settings. Other details can be found in the Appendix.

Similar to the well-specified setting, we also plot the running proportion of arms over time and the most frequently selected arm trajectories in Figure 13 (a) and (b) respectively in Appendix E.

Although the arm selections remain mostly similar as in the well-specified settings, we do notice that UCB tends to explore more in the misspecified setting (with the same exploration parameter $\alpha = 1$ as in the well-specified setting), which also results in better parameter estimates for UCB as is seen in Figure 14 in Appendix E.

These findings underline a pragmatic principle for on-farm experimentation: when the *shape* of the assumed response is qualitatively correct, e.g. monotone-increasing with a plateau, model-based bandits remain sample efficient even if the functional form is wrong. Conversely, generic linear or purely non-parametric methods may demand far more data before converging, a luxury seldom available to smallholders. Hence, coupling modest domain knowledge with sequential decision-making promises a robust path toward Sustainable Development Goal 2 by enabling data-scarce producers to improve input efficiency without costly large-scale trials.

6 Real data analysis: Optimizing nitrogen rates for corn in the U.S. Midwest

We evaluate the proposed bandit algorithms on publicly available multi-site corn nitrogen field-trial data from a public–industry partnership spanning the 2014–2016 growing seasons, collected under standardized protocols across U.S. Midwest institutions (Ransom et al., 2021). For reproducibility, we construct an analysis-ready subset by retaining only the variables needed for sequential decision-making (state, site, year, block, nitrogen rate, and yield), removing incomplete records, and harmonizing formats; the processed files are provided in the supplement and variable descriptions in Table 7 of the Appendix.

6.1 Offline replay protocol and profit objective

The goal is to assess whether bandit-style decision rules can adaptively choose nitrogen rates to improve an economic objective. Because the trials were not run sequentially for online learning, we evaluate policies using an *offline replay* procedure: we treat each *round* as a grouped set of plots sharing the same environment and management conditions, and within that round the available actions are the nitrogen rates observed in that group. These rounds should be interpreted as *decision instances* rather than a true temporal sequence.

Two complementary regimes. We report results in two settings that mirror data availability in practice: (i) a data-limited case study at a single site, and (ii) a pooled multi-site evaluation that aggregates across similar environments. Specifically:

1. **Single-site case study (short horizon):** Urbana, IL (2014–2016), with rounds defined as (Year, Block). With blocks $\in \{1, 2, 3, 4\}$ and years $\in \{2014, 2015, 2016\}$, this yields $T = 12$ rounds. In other words, we fix State = IL and Site = Urbana, and treat Block $\in \{1, 2, 3, 4\}$ as geographically similar sub-units within the site observed over Year $\in \{2014, 2015, 2016\}$, giving $T = 3 \times 4 = 12$ rounds.
2. **Pooled low-productivity evaluation (longer horizon):** We pool all observations labeled Site_Prod = low, where Site_Prod is the dataset’s within-state productivity label indicating that a site-year falls into the lower-yielding group relative to other sites in the same state (based on the study’s baseline/expected-yield information). Rounds are defined as (State, Site, Year, Block), yielding a modestly larger horizon (e.g., $T \approx 56$ in our filtered sample; see Table 8 in the Appendix for locations and counts). Restricting to the

low-productivity subset reduces cross-round heterogeneity so that a single *non-contextual* policy is meaningful; otherwise, the profit-maximizing nitrogen rate can vary substantially across environments and would naturally call for a contextual (covariate-dependent) model.

Profit objective. In both regimes we optimize $\Pi(x) = p_y Y(x) - p_x x$, where $Y(x)$ is yield (bu/ac) at nitrogen rate x (lb N/ac), p_y is the corn price (\$/bu), and p_x is the nitrogen cost (\$/lb N). We use year-specific corn prices from publicly available summary data (USDA National Agricultural Statistics Service, 2017) and compute p_x from December Midwest urea prices using urea’s 46% N analysis (Yara North America, Inc., 2018); the resulting year-by-year values are summarized in Table 9 in the Appendix.

Bandit feedback, oracle benchmark, and regret. Within each round, multiple nitrogen-rate treatments are observed (with replicated plots), but the policy selects a single rate x_t and the replay reveals only the corresponding realized outcome (mean over replicates at x_t , i.e., mean-reward mode), thereby emulating bandit feedback. The underlying experiment that generated the original data uses a randomized complete block design (RCBD), so outcomes for all treatments are available within each round for evaluation (Ransom et al., 2021). We define the within-round oracle action (best mean profit among the observed rates in that round)

$$x_t^* \in \arg \max_{x \in \mathcal{X}_t} \bar{\Pi}_t(x),$$

where \mathcal{X}_t is the set of observed nitrogen rates in round t and $\bar{\Pi}_t(x)$ is mean profit across replicates at rate x . We then compute per-round regret as $r_t = \bar{\Pi}_t(x_t^*) - \Pi_t(x_t)$ (unit in \$/ac) and cumulative regret $R_t = \sum_{s=1}^t r_s$ (units in \$/ac), reporting both R_t and R_t/t .

To reduce sensitivity to a single ordering, we repeat the offline replay over 100 randomized orderings; for the pooled analysis we sort rounds by year and shuffle within year, then average curves across replays.

6.2 Results

Figure 7(a)-(d) summarize offline replay performance in terms of *cumulative regret* $R_t = \sum_{s=1}^t r_s$ and *average regret* R_t/t , where $r_t = \bar{\Pi}_t(x_t^*) - \Pi_t(x_t)$ is the within-round profit gap to the oracle benchmark as defined above. We report mean trajectories over 100 replay replications, with pointwise 95% confidence bands.

Algorithms compared. In the single-site case study, we compare the three nonlinear model-based strategies with profit maximization objective, i.e., model-based ϵ -greedy (Alg 5), nonlinear-UCB (Alg. 6), and ViOlin (Alg. 7), and nonparametric k-NN UCB (Alg. 9) against a random baseline that selects a nitrogen rate uniformly at random from the rates available in each round. For the pooled low-productivity analysis, we additionally include LinUCB (Alg. 8) and kNN-UCB (Alg. 9) as parametric and nonparametric benchmarks, respectively. In all model-based methods, the fitted reward model is the quadratic-plateau response with the profit objective $\Pi(x) = p_y Y(x) - p_x x$ (Section 2.1), and regret is measured in profit units (\$/ac).

Urbana, IL (short horizon; $T = 12$ rounds). In the data-limited Urbana, IL case study (rounds defined by year \times block), the nonlinear model-based algorithms consistently reduce regret relative to the random baseline (Figure 7(a)-(b)). Both the cumulative regret curves (left) and the corresponding average regret R_t/t curves (right) show that leveraging agronomically motivated

nonlinear structure can be substantially more sample-efficient when only a handful of decisions are available.

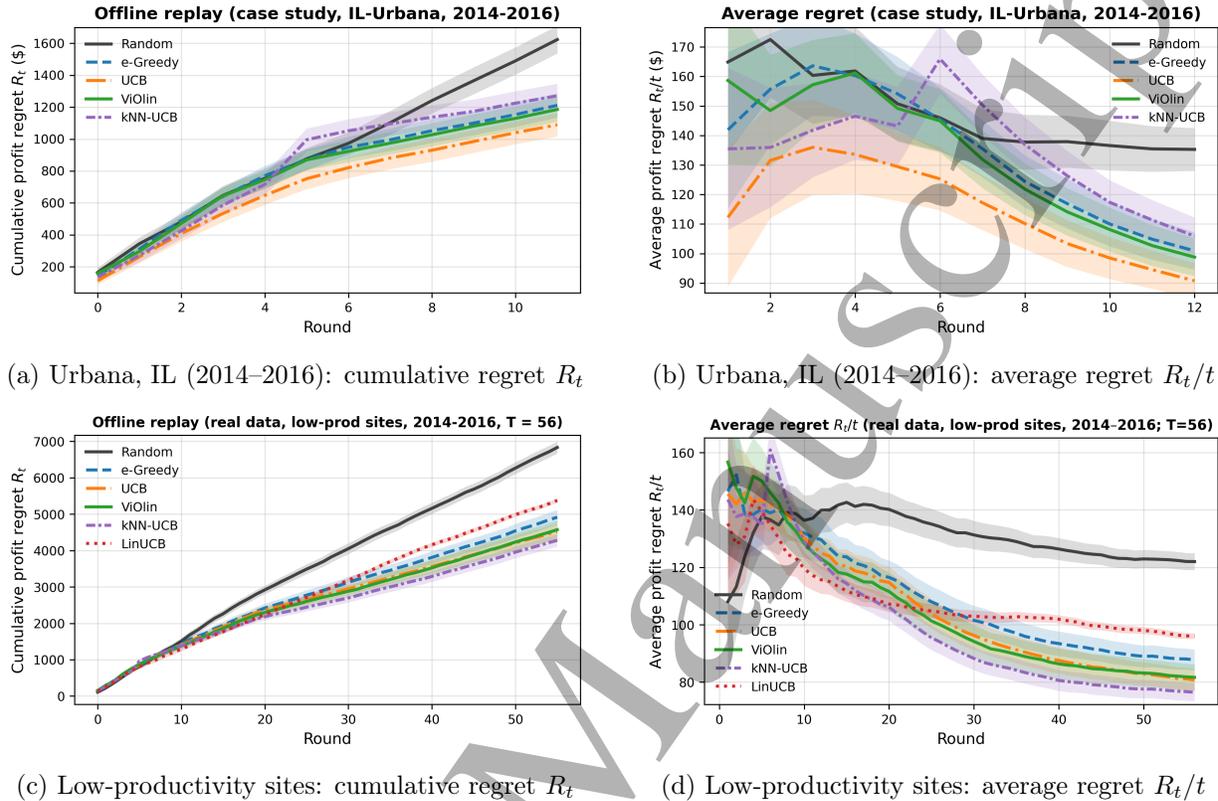


Figure 7: **Offline replay on real multi-site corn nitrogen trials (2014–2016): profit regret under a quadratic-plateau model.** Reward is per-acre profit $\Pi(x) = p_y Y(x) - p_x x$, where p_y is the corn price (USDA–NASS Crop Values) and p_x is the nitrogen cost as in Table 9. Panels (a)–(b) show the data-limited Urbana, IL case study with rounds defined by (Year, Block) ($T = 12$). Panels (c)–(d) show the pooled low-productivity evaluation with rounds defined by (State, Site, Year, Block) (longer horizon). Curves are means over replay replications; shaded bands are pointwise 95% confidence intervals.

Low-productivity sites across states (longer horizon; $T \approx 56$ rounds). When pooling all low-productivity site-years across states (rounds defined by state \times site \times year \times block), we observe a different pattern (Figure 7(c)-(d)). After an initial transient, the nonparametric kNN-UCB becomes increasingly competitive and can outperform the parametric baselines, while LinUCB degrades under cross-site heterogeneity. This is consistent with the fact that a single global linear approximation is poorly suited to multi-environment (different locations) response variability, whereas local smoothing in action space can provide robustness once sufficient history accumulates.

Taken together, these results mirror the main message from the simulations while highlighting a practical distinction between regimes: under short horizons, nonlinear model-based bandits provide clear benefits over uninformed baselines, whereas under longer horizons with heterogeneous pooled environments, flexible nonparametric benchmarks can become competitive. Extending to mixed productivity regimes is natural future work using contextual bandits with site and soil covariates.

7 Conclusion and discussion

We developed and evaluated nonlinear, model-based bandit algorithms for sequential fertilizer decision-making using agronomy-standard mechanistic yield–nitrogen response families (Mitscherlich, quadratic-plateau, Michaelis–Menten, and logistic). By framing nitrogen-rate selection as an online learning problem, these models become interpretable decision rules that update as data accrue. Our primary objective is economic: we optimize profit (revenue minus fertilizer cost) rather than yield alone, so recommendations directly reflect realistic input-cost trade-offs. To our knowledge, our study provides one of the first systematic comparisons in the data-limited regime demonstrating when agronomy-standard nonlinear response modeling embedded in bandit decision rules yields clear advantages over linear and model-free baselines for profit-oriented nitrogen recommendations made sequentially across seasons.

In simulations, nonlinear model-based bandits achieved substantially lower profit regret than linear and nonparametric baselines in well-specified settings, highlighting the value of incorporating domain structure in small-sample regimes. Under misspecification, performance degrades as expected, but mechanistic models remain competitive when they provide a reasonable approximation and data are limited. We also included an offline replay case study on publicly available multi-site corn nitrogen trials (Ransom et al., 2021), which corroborates the practical message: in short horizons, nonlinear model-based strategies can be more sample-efficient than uninformed baselines, while in longer pooled (multiple locations) regimes, nonparametric methods can become competitive.

Beyond average performance, it is important to characterize the uncertainty surrounding recommended nitrogen rates, particularly given the inherent complexity of agroecosystems. Our nonlinear model-based framework provides ingredients for such summaries through parameter covariance estimates, which can be propagated to quantify uncertainty in predicted profit and in the profit-maximizing rate. Rather than reporting only a point recommendation, future implementations could present near-optimal ranges of nitrogen rates whose predicted profits are statistically indistinguishable from the maximum, highlighting when the optimum is sharply identified versus when modest rate adjustments carry limited economic risk (Palmero et al., 2026). Developing decision-oriented uncertainty reporting and risk-aware extensions, such as, coherent risk measures (e.g., CVaR) or posterior sampling strategies (e.g., Thompson sampling) within nonlinear bandit frameworks is therefore an important direction for future work.

Several other directions merit further work: (i) principled model selection/averaging within the bandit loop to improve robustness to misspecification; (ii) extensions to spatially varying or multi-dimensional decisions (e.g., variable-rate application); (iii) contextual and nonstationary formulations that incorporate soil/weather covariates and season-to-season shifts; and (iv) tighter regret guarantees under structured misspecification and extensions to delayed or partially observed outcomes.

Reproducibility statement

All simulations were implemented in Python 3.11 using standard scientific libraries (NumPy, SciPy, Matplotlib). We report the data-generating processes, parameter values, and hyperparameters in full within the paper and Appendix, including initial values used for nonlinear least-squares estimation, fertilizer and grain price settings, and noise levels. Each experiment was repeated independently across 10 replicates with fixed randomization schemes to assess variability. The code to reproduce all simulation experiments and figures, together with the analysis-ready files

used for the offline real-data replay evaluation, are provided in the Supplementary Material and will be made publicly available on GitHub upon publication.

Data availability statement

No new data were collected for this study. The real-data analysis uses the publicly available multi-site corn nitrogen trial dataset of Ransom et al. (2021).

Ethics statement

This study does not involve human participants, human data, or animals, and therefore ethics approval was not required.

Funding statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24.
- Abdipourchenarestansofla, M. and Piepho, H.-P. (2025). In season estimation of economic optimum nitrogen rate with remote sensing multispectral indices and historical telematics field-operation data. *Precision Agriculture*, 26(2):34.
- Agarwal, A., Foster, D. P., Hsu, D., Kakade, S. M., and Langford, J. (2012). Contextual bandit learning with predictable rewards. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 22, pages 19–26.
- Akhavizadegan, F., Ansarifard, J., Wang, L., Huber, I., and Archontoulis, S. V. (2021). A time-dependent parameter estimation framework for crop modeling. *Scientific Reports*, 11(1):11437.
- Arya, S. and Yang, Y. (2020). Randomized allocation with nonparametric estimation for contextual multi-armed bandits with delayed rewards. *Statistics & Probability Letters*, page 108818.
- Auer, P. and Ortner, R. (2010). UCB revisited: Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica*, 61(1-2):55–65.
- Bastani, H., Bayati, M., and Khosravi, K. (2020). Mostly exploration-free algorithms for contextual bandits. *Management Science*, 66(2):780–803.

- 1
2
3 Baudry, D., Gautron, R., Kaufmann, E., and Maillard, O. (2021). Optimal Thompson sampling
4 strategies for support-aware cvar bandits. In *International Conference on Machine Learning*,
5 pages 716–726. PMLR.
6
7 B elanger, G., Walsh, J. R., Richards, J. E., Milburn, P. H., and Ziadi, N. (2000). Comparison
8 of three statistical models describing potato yield response to nitrogen fertilizer. *Agronomy*
9 *Journal*, 92(5):902–908.
10
11 Berry, D. A. and Fristedt, B. (1985). Bandit problems: sequential allocation of experiments
12 (monographs on statistics and applied probability). *London: Chapman and Hall*, 5:71–87.
13
14 Chu, W., Li, L., Reyzin, L., and Schapire, R. E. (2011). Contextual bandits with linear payoff
15 functions. In *Proceedings of the 14th International Conference on Artificial Intelligence and*
16 *Statistics (AISTATS)*, volume 15, pages 208–214.
17
18 Dani, V., Hayes, T. P., and Kakade, S. M. (2008). Stochastic linear optimization under bandit
19 feedback. In *Conference on Learning Theory (COLT)*, pages 355–366.
20
21 De Bruin, J. L. and Pedersen, P. (2008). Soybean seed yield response to planting date and seeding
22 rate in the upper midwest. *Agronomy Journal*, 100(3):696–703.
23
24 Dhanoa, M. S., Sanderson, R., Cardenas, L., Shepherd, A., Chadwick, D., Powell, C., Ellis, J.,
25 L opez, S., and France, J. (2022). Overview and application of the Mitscherlich equation and
26 its extensions to estimate the soil nitrogen pool fraction associated with crop yield and Nitrous
27 Oxide emission. *Advances in Agronomy*, 174:269–295.
28
29 Dobermann, A., Bruulsema, T., Cakmak, I., Gerard, B., Majumdar, K., McLaughlin, M., Reidsma,
30 P., Vanlauwe, B., Wollenberg, L., Zhang, F., et al. (2022). Responsible plant nutrition: A new
31 paradigm to support food system transformation. *Global Food Security*, 33:100636.
32
33 Dong, K., Yang, J., and Ma, T. (2021). Provable model-based nonlinear bandit and reinforce-
34 ment learning: Shelve optimism, embrace virtual curvature. *Advances in Neural Information*
35 *Processing Systems*, 34:26168–26182.
36
37 Filippi, S., Cappe, O., Garivier, A., and Szepesv ari, C. (2010). Parametric bandits: The generalized
38 linear case. In *Advances in Neural Information Processing Systems*, volume 23, pages 586–594.
39
40 Foster, D. J., Krishnamurthy, A., Syrgkanis, V., and Agarwal, A. (2018). Practical contextual
41 bandits with regression oracles. In *Proceedings of the 35th International Conference on Machine*
42 *Learning (ICML)*, pages 1539–1548.
43
44 Gautron, R., Maillard, O.-A., Preux, P., Corbeels, M., and Sabbadin, R. (2022). Reinforcement
45 learning for crop management support: Review, prospects and challenges. *Computers and Elec-*
46 *tronics in Agriculture*, 200:107182.
47
48 Ghosh, A., Chowdhury, S. R., and Gopalan, A. (2017). Misspecified linear bandits. In *Proceedings*
49 *of the AAAI Conference on Artificial Intelligence*, volume 31.
50
51 Greenewald, K., Krishnamurthy, A., Lattimore, T., Slivkins, A., and Syrgkanis, V. (2021). Adapt-
52 ing to misspecification in contextual bandits. In *Proceedings of the 38th International Conference*
53 *on Machine Learning (ICML)*, pages 3832–3842.
54
55
56
57
58
59
60

- 1
2
3 Hochman, Z. and Carberry, P. (2011). Emerging consensus on desirable characteristics of tools to
4 support farmers' management of climate risk in Australia. *Agricultural Systems*, 104(6):441–450.
5
6 Huang, Q. (2025). Precision agriculture optimization based on multi-armed bandits algorithm:
7 Wheat yield optimization under different temperature and precipitation conditions. In *ITM*
8 *Web of Conferences*, volume 73, page 01013. EDP Sciences.
9
10 Jabed, M. A. and Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive
11 review of machine learning and deep learning approaches, with insights for future research and
12 sustainability. *Heliyon*.
13
14 Khairunniza-Bejo, S., Mustaffha, S., and Ismail, W. I. W. (2014). Application of artificial neural
15 network in predicting crop yield: A review. *Journal of Food Science and Engineering*, 4(1):1.
16
17 Khaki, S., Wang, L., and Archontoulis, S. V. (2020). A CNN-RNN framework for crop yield
18 prediction. *Frontiers in Plant Science*, 10:1750.
19
20 Langford, J. and Zhang, T. (2008). The epoch-greedy algorithm for multi-armed bandits with side
21 information. In *Advances in Neural Information Processing Systems*, pages 817–824.
22
23 Lattimore, T., Munos, R., and Szepesvári, C. (2020). Learning with misspecified linear bandit
24 models. In *Conference on Learning Theory*, pages 1112–1146.
25
26 Lattimore, T. and Szepesvári, C. (2018). Bandit algorithms. *Cambridge University Press*.
27
28 Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to
29 personalized news article recommendation. In *Proceedings of the 19th International Conference*
30 *on World Wide Web*, pages 661–670. ACM.
31
32 Lopez, S., France, J., Gerrits, W., Dhanoa, M., Humphries, D., and Dijkstra, J. (2000). A general-
33 ized Michaelis-Menten equation for the analysis of growth. *Journal of animal science*, 78(7):1816–
34 1828.
35
36 Lu, Y., Xu, Z., and Tewari, A. (2021). Bandit algorithms for precision medicine. *arXiv preprint*
37 *arXiv:2108.04782*.
38
39 Matavel, C. E., Meyer-Aurich, A., and Piepho, H.-P. (2025). Bayesian-optimized experimen-
40 tal designs for estimating the economic optimum nitrogen rate: a model-averaging approach.
41 *Agronomy Journal*, 117(3):e70087.
42
43 Mia, M. S., Habibi, L. N., Matsui, T., and Tanaka, T. S. T. (2024). A Bayesian approach to
44 assessing uncertainty in the effect of fertilization strategies on paddy rice yield via multiple
45 on-farm experiments. *Plant Production Science*, 27(3):197–211.
46
47 Miguez, F., Archontoulis, S., and Dokoochaki, H. (2018). Nonlinear regression models and applica-
48 tions. *Applied statistics in agricultural, biological, and environmental sciences*, pages 401–447.
49
50 Ngómbe, J. N. and Brorsen, B. W. (2022). Bayesian optimal dynamic sampling procedures for
51 on-farm field experimentation. *Precision Agriculture*, 23(6):2289–2311.
52
53 Osband, I. and Van Roy, B. (2014). Model-based reinforcement learning and the Eluder dimension.
54 *Advances in Neural Information Processing Systems*, 27.
55
56
57
58
59
60

- 1
2
3 Palmero, F., Davidson, E. A., Guan, K., Eagle, A. J., Birgé, H. E., Prasad, P. V., Hefley, T. J.,
4 Schussler, J. R., and Ciampitti, I. A. (2026). Environmental and societal costs of maize produc-
5 tion decrease by addressing the uncertainty in nitrogen rate recommendations. *Nature Commu-*
6 *nications*.
- 8 Poursina, D. and Brorsen, B. W. (2024). Fully Bayesian economically optimal design for a spatially
9 varying coefficient linear stochastic plateau model over multiple years. *Stochastic Environmental*
10 *Research and Risk Assessment*, 38(3):1089–1098.
- 12 Rakhlin, A., Sridharan, K., and Tewari, A. (2015a). Online learning via sequential complexities.
13 *Journal of Machine Learning Research*, 16(6):155–186.
- 15 Rakhlin, A., Sridharan, K., and Tewari, A. (2015b). Sequential complexities and uniform martin-
16 gale laws of large numbers. *Probability theory and related fields*, 161:111–153.
- 18 Ransom, C. J., Clark, J., Bean, G. M., Bandura, C., Shafer, M. E., Kitchen, N. R., Camberato,
19 J. J., Carter, P. R., Ferguson, R. B., Fernández, F., Franzen, D. W., Laboski, C. A., Myers,
20 D. B., Nafziger, E. D., Sawyer, J. E., and Shanahan, J. (2021). Data from a public, industry
21 partnership for enhancing corn nitrogen research. *Agronomy Journal*, 113(5):4429–4436.
- 23 Ransom, C. J., Kitchen, N. R., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernández, F. G.,
24 Franzen, D. W., Laboski, C. A. M., Nafziger, E. D., Sawyer, J. E., Scharf, P. C., and Shanahan,
25 J. F. (2020). Corn nitrogen rate recommendation tools’ performance across eight US midwest
26 corn belt states. *Agronomy Journal*, 112(1):470–492.
- 28 Reeve, H., Mellor, J., and Brown, G. (2018). The k-nearest neighbour UCB algorithm for multi-
29 armed bandits with covariates. In *Algorithmic Learning Theory*, pages 725–752. PMLR.
- 31 Rigollet, P. and Zeevi, A. (2010). Nonparametric bandits with covariates. *COLT 2010*, page 54.
- 33 Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American*
34 *Mathematical Society*, 58(5):527–535.
- 36 Saikai, Y., Mitchell, P. D., and Colquhoun, J. (2018). Multi-armed bandit for experimental plot
37 selection. Technical report, Agricultural and Applied Economics Association.
- 39 Saikai, Y., Patel, V., and Mitchell, P. D. (2020). Machine learning for optimizing complex site-
40 specific management. *Computers and Electronics in Agriculture*, 174:105381.
- 42 Sepaskhah, A. R., Fahandezh-Saadi, S., and Zand-Parsa, S. (2011). Logistic model application for
43 prediction of maize yield under water and nitrogen management. *Agricultural Water Manage-*
44 *ment*, 99(1):51–57.
- 46 Shahhosseini, M., Martinez-Feria, R. A., Hu, G., and Archontoulis, S. V. (2019). Maize yield
47 and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*,
48 14(12):124026.
- 50 Sliwkins, A. (2014). Contextual bandits with similarity information. *Journal of Machine Learning*
51 *Research*, 15:2533–2568.
- 53 Smith, M. J. (2018). Getting value from artificial intelligence in agriculture. *Animal Production*
54 *Science*, 60(1):46–54.

- 1
2
3 Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. (2010). Gaussian process optimization
4 in the bandit setting: No regret and experimental design. In *Proceedings of the International*
5 *Conference on Machine Learning, 2010*.
6
7 Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. (2012). Information-theoretic re-
8 gret bounds for Gaussian process optimization in the bandit setting. *IEEE transactions on*
9 *information theory*, 58(5):3250–3265.
10
11 Sun, W., Jiang, N., Krishnamurthy, A., Agarwal, A., and Langford, J. (2019). Model-based RL
12 in contextual decision processes: Pac bounds and exponential improvements over model-free
13 approaches. In *Conference on learning theory*, pages 2898–2933. PMLR.
14
15 Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002). Agricultural
16 sustainability and intensive production practices. *Nature*, 418(6898):671–677.
17
18 USDA National Agricultural Statistics Service (2017). Crop values 2016 summary. Annual Sum-
19 mary CpV1 2-1 (17), U.S. Department of Agriculture, National Agricultural Statistics Service.
20 Accessed 2026-01-19.
21
22 Valko, M., Korda, N., Munos, R., Flaounas, I., and Cristianini, N. (2013). Finite-time analysis of
23 kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty*
24 *in Artificial Intelligence*, pages 654–663.
25
26 Wang, K., Kallus, N., and Sun, W. (2023). Near-minimax-optimal risk-sensitive reinforcement
27 learning with CVaR. In *International Conference on Machine Learning*, pages 35864–35907.
28 PMLR.
29
30 Yang, Y. and Zhu, D. (2002). Randomized allocation with nonparametric estimation for a multi-
31 armed bandit problem with covariates. *The Annals of Statistics*, (1):100–121.
32
33 Yara North America, Inc. (2018). Yaravera urea granular 46-0-0: Guaranteed analysis (bulk label).
34 Technical report, Yara North America, Inc. Confirms urea is 46% nitrogen by weight (46-0-0).
35 Accessed 2026-01-19.
36
37 Zhou, D., Li, L., and Gu, Q. (2020). Neural contextual bandits with UCB-based exploration. In
38 *International Conference on Machine Learning*, pages 11492–11502. PMLR.
39
40
41

42 Appendix

43 A Additional Background on expressive parametric function classes.

44 In this section, we discuss the more expressive Neural bandits and provide regret summaries for
45 bandit algorithms pertaining to other parametric model classes as discussed in Section 4.
46
47
48
49

50 **Neural Bandits** Recent work has extended bandit algorithms to highly expressive nonlinear
51 function classes via neural networks. For instance, NeuralUCB (Zhou et al., 2020) achieves regret
52 rates of order
53

$$54 R_T = \tilde{O}\left(\tilde{d}(\lambda)\sqrt{T}\right),$$

where $\tilde{d}(\lambda)$ is the effective dimension of the neural tangent kernel associated with the underlying network. For example, for a two-layer ReLU neural network with input dimension d and width m , it holds that $\tilde{d}(\lambda) \leq d \text{poly}(m, \log(T/\lambda))$, so that the regret rate can be written as

$$R_T = \tilde{O}\left(d \text{poly}(m) \sqrt{T}\right).$$

Here, $\text{poly}(m, \cdot)$ denotes a polynomial function of the network width and the indicated arguments. Thus, the regret bound increases with both the expressiveness of the neural network (through m) and the ambient input dimension d , but retains the \sqrt{T} scaling characteristic of parametric (linear) bandit models. However, these improvements come with heavy computational costs, the need for large data, and limited interpretability.

Table 3: Typical regret rates for common stochastic bandit models (for context). Here T is the number of decision rounds (the “time horizon”) and d is the number of covariates (features). The notation $\tilde{O}(\cdot)$ hides logarithmic factors. The NeuralUCB entry depends on the neural-network width m and other architecture constants; see the cited reference for details.

Model Class	Regret Bound	Reference
Linear Bandit	$O(d\sqrt{T} \log T)$	Abbasi-Yadkori et al. (2011)
GLM Bandit	$\tilde{O}(d\sqrt{T})$	Filippi et al. (2010)
Nonparametric (Lipschitz, RKHS)	$\tilde{O}(T^{\frac{d+1}{d+2}})$	Slivkins (2014)
NeuralUCB (2-layer NN)	$\tilde{O}(d \text{poly}(m) \sqrt{T})$	Zhou et al. (2020)

B Rademacher complexity for parametric nonlinear function classes

Proof of Theorem 2. We follow the approach of Rakhlin et al. (2015b). Let $N_1(\alpha, \mathcal{F}, T)$ denote the sequential covering number in the ℓ_1 -norm at scale α . From equation (9) of their paper, we have:

$$\mathfrak{R}_T(\mathcal{F}) \leq \alpha + \sqrt{\frac{2 \log N_1(\alpha, \mathcal{F}, T)}{T}}. \quad (9)$$

Using Corollary 1 from the same paper, the ℓ_∞ -sequential covering number satisfies:

$$N_\infty(\alpha, \mathcal{F}, T) \leq \left(\frac{2eT}{\alpha}\right)^{\text{fat}_\alpha(\mathcal{F})},$$

and since $N_1(\alpha, \mathcal{F}, T) \leq N_\infty(\alpha, \mathcal{F}, T)$, we obtain:

$$\log N_1(\alpha, \mathcal{F}, T) \leq \text{fat}_\alpha(\mathcal{F}) \cdot \log\left(\frac{2eT}{\alpha}\right). \quad (10)$$

For any class $\mathcal{F} \subseteq [-B_{\mathcal{F}}, B_{\mathcal{F}}]^{\mathcal{X}}$, we have from Theorem 10 of Rakhlin et al. (2015a):

$$\text{fat}_\alpha(\mathcal{F}) \leq \left\lceil \frac{2B_{\mathcal{F}}}{\alpha} \right\rceil. \quad (11)$$

Now, using the bounds in (10) and (11) in (9) and choosing $\alpha = \frac{B_{\mathcal{F}}}{\sqrt{T}}$, we get:

$$\left\lceil \frac{2B_{\mathcal{F}}}{\alpha} \right\rceil = \lceil 2\sqrt{T} \rceil \leq 3\sqrt{T}, \quad \log\left(\frac{2eT}{\alpha}\right) = \log\left(\frac{2eT^{3/2}}{B_{\mathcal{F}}}\right) = \mathcal{O}(\log T).$$

Thus,

$$\mathfrak{R}_T(\mathcal{F}) \leq \frac{B_{\mathcal{F}}}{\sqrt{T}} + \sqrt{\frac{6\sqrt{T} \cdot \log T}{T}} = \frac{B_{\mathcal{F}}}{\sqrt{T}} + \sqrt{\frac{6 \log T}{\sqrt{T}}}.$$

The dominant term is $\frac{B_{\mathcal{F}}\sqrt{\log T}}{\sqrt{T}}$, so we conclude:

$$\mathfrak{R}_T(\mathcal{F}) \leq C \cdot B_{\mathcal{F}} \cdot \sqrt{\frac{\log T}{T}}$$

for some universal constant C . □

B.1 Computation of the model-dependent bound $B_{\mathcal{F}}$

Figure 1 visualizes the upper bound in Theorem 2 using a model-dependent magnitude bound $B_{\mathcal{F}}$. For each yield-response family \mathcal{F} , we define

$$B_{\mathcal{F}} := \sup_{x \in [0, 250]} |f(x; \theta)|,$$

where the domain $[0, 250]$ matches the fertilizer range used in the simulation study (Section 5) and $f(\cdot; \theta)$ is evaluated under the parameter settings used in that section.

Example: quadratic-plateau. For the quadratic-plateau model

$$Y(x) = \begin{cases} a + bx + cx^2, & x \leq x_0, \\ a + bx_0 + cx_0^2, & x > x_0, \end{cases}$$

with $(a, b, c, x_0) = (80, 1.2, -0.003, 180)$ and $c < 0$, the quadratic component is concave on $[0, x_0]$ and the function is constant for $x > x_0$. Hence the maximum on $[0, 250]$ occurs at $x = x_0$, and

$$B_{\text{QP}} = \sup_{x \in [0, 250]} f(x) = f(x_0) = 80 + 1.2(180) - 0.003(180)^2 = 198.8 \text{ bu/ac}.$$

Values used in Figure 1. For the other model families (Mitscherlich, Logistic, and Michaelis–Menten), $f(x)$ is monotone increasing on $[0, 250]$ under the parameters used in Section 5, so $B_{\mathcal{F}}$ is attained at $x = 250$. Table 4 lists the resulting values.

Table 4: Model-dependent bounds $B_{\mathcal{F}} = \sup_{x \in [0, 250]} |f(x)|$ (bu/ac) used in Figure 1, computed under the parameter settings in Section 5.

Model family \mathcal{F}	$B_{\mathcal{F}}$ (bu/ac)
Quadratic-plateau	198.80
Mitscherlich	197.18
Logistic	189.77
Michaelis–Menten	167.14

C Finding optimal profit maximizing arms for different non-linear models

Recall that $\Pi(x)$ expresses the trade-off between revenue from yield and the cost of input. To find the economically optimal fertilizer dose x^* , we differentiate the profit function with respect to x : Below, we derive x^* for each model considered in this work.

Mitscherlich Model: The Mitscherlich yield response is

$$Y(x) = A(1 - e^{-bx}),$$

so the profit function is

$$\Pi(x) = p_y A(1 - e^{-bx}) - p_x x.$$

The optimal dose x^* solves

$$\frac{d\Pi}{dx} = p_y A b e^{-bx} - p_x = 0,$$

giving

$$x^* = -\frac{1}{b} \ln\left(\frac{p_x}{p_y A b}\right).$$

Note: If $\frac{p_x}{p_y A b} \geq 1$, then $x^* = 0$.

Quadratic Model with Threshold Suppose

$$Y(x) = \begin{cases} a + bx + cx^2, & x \leq x_0 \\ a + bx_0 + cx_0^2, & x > x_0 \end{cases}$$

with threshold x_0 . Then

$$\Pi(x) = p_y Y(x) - p_x x.$$

For $x \leq x_0$, set the derivative to zero:

$$\frac{d\Pi}{dx} = p_y(b + 2cx) - p_x = 0 \implies x^* = \frac{1}{2c} \left(\frac{p_x}{p_y} - b \right).$$

Thus,

$$x^* = \min \left\{ x_0, \max \left\{ 0, \frac{1}{2c} \left(\frac{p_x}{p_y} - b \right) \right\} \right\}.$$

If $x > x_0$, profit decreases with increasing x .

Michaelis-Menten (MM) Model With

$$Y(x) = \frac{ax}{b + x},$$

the profit is

$$\Pi(x) = p_y \frac{ax}{b + x} - p_x x.$$

Setting the derivative to zero,

$$\frac{d\Pi}{dx} = p_y \frac{ab}{(b + x)^2} - p_x = 0 \implies (b + x)^2 = \frac{abp_y}{p_x}$$

$$x^* = \sqrt{\frac{abp_y}{p_x}} - b.$$

We set $x^* = 0$ if the right side is negative.

Logistic Model For the logistic response

$$Y(x) = \frac{A}{1 + e^{-B(x-C)}},$$

profit is

$$\Pi(x) = p_y \frac{A}{1 + e^{-B(x-C)}} - p_x x.$$

Let $u = e^{-B(x-C)}$, so $x = C - \frac{1}{B} \ln u$. Setting the derivative to zero,

$$p_y A \frac{Bu}{(1+u)^2} = p_x,$$

which yields a quadratic in u :

$$u^2 + \left(2 - \frac{Bp_y A}{p_x}\right)u + 1 = 0.$$

Let

$$\gamma = \frac{Bp_y A}{p_x}.$$

The positive root is

$$u^* = \frac{\gamma - 2 - \sqrt{(\gamma - 2)^2 - 4}}{2},$$

and the optimal dose is

$$x^* = C - \frac{1}{B} \ln u^*.$$

If $u^* \leq 0$ or $x^* < 0$, set $x^* = 0$.

D Baseline Algorithm descriptions

We compare the non-linear model based algorithms to the nonparametric and linear baselines. Specifically, we compare these with the LinUCB and k -Nearest Neighbor UCB algorithm. For the sake of completeness, we describe these algorithms below.

The **LinUCB algorithm** (Algorithm 8) is a classic model-based bandit method that assumes a linear relationship between the reward and the action (or its features). At each round, LinUCB fits a linear regression model to the observed data (line 7) and predicts the expected yield for each fertilizer rate (line 8a). It computes an upper confidence bound (UCB) for each action by adding a model-based uncertainty term, proportional to the standard error of the prediction, to the profit estimate (lines 8b-d). The fertilizer rate with the highest UCB is selected for the next trial (line 9), balancing exploration of uncertain arms and exploitation of high-yield arms. While LinUCB is efficient and easy to implement, its performance can degrade if the true reward function is nonlinear or misspecified.

The **kNN-UCB algorithm** (Algorithm 9) is a nonparametric bandit method that does not assume a specific model for the yield response. Instead, for each candidate fertilizer rate, it estimates the expected yield by averaging the observed yields of the k most similar (nearest) rates previously tried (line 8). The uncertainty in this prediction is quantified by the sample variance among these k neighbors (line 9). As with LinUCB, an upper confidence bound is constructed for each candidate

rate (line 11), and the rate with the highest UCB is selected (line 13). Although kNN-UCB is robust to model misspecification and can capture complex nonlinearities, it generally requires more data to achieve accurate estimates, particularly when the action space is large or sparsely explored.

D.1 Toy Illustration of Algorithm Behavior

To make the exploration–exploitation logic more tangible, we provide a simple illustration using a Mitscherlich yield function

$$Y(x) = 80 + 120(1 - e^{-0.015x}),$$

with crop price $p_y = \$5$ per unit yield (bu^{-1}), fertilizer cost $p_x = \$0.7$ per unit ($\text{lb}^{-1} \text{N}$), and action grid $\{0, 50, 100, 150, 200\}$ lb N /ac. Table 5 shows four rounds of decision making under three algorithms: ϵ -Greedy, UCB, and ViOlin.

Table 5: Toy demonstration of algorithm behavior (profit in \$).

Round	ϵ -Greedy		UCB		ViOlin	
	Action	Profit	Action	Profit	Action	Profit
1	Explore 50	541	Explore 100	610	Explore 50	540
2	Explore 150	650	Explore 150	648	Exploit 150	649
3	Exploit 150	648	Exploit 150	647	Exploit 150	648
4	Explore 50	541	Explore 200	645	Exploit 150	647

This toy comparison highlights the distinct decision logics:

- ϵ -Greedy alternates between random exploration and exploiting the currently estimated best dose.
- UCB incorporates an explicit uncertainty bonus, occasionally sampling higher but uncertain doses (round 4).
- ViOlin emphasizes curvature and stability, quickly converging to the plateau dose and avoiding unnecessary exploration.

Even in this simple 4-round setting, the three strategies exhibit clearly different exploration styles. Such illustrations help convey to applied readers that algorithm choice affects not only long-run performance, but also the sequence of recommendations farmers may observe in practice.

E Additional Simulation Results

In this section, we present the simulation results for other non-linear models that we considered in the simulation experiment of Section 5. First for the well-specified setting, we present the results for the Logistic and Mitscherlich model that was fit instead of quadratic plateau model. For reproducibility, Table 6 reports the true parameter values used to generate data in our simulations, alongside the initial values supplied to the nonlinear least-squares estimation routines. These initialization choices were selected to be reasonably close to the true parameters, reflecting what might be obtained from domain knowledge or prior agronomic experiments. As discussed in Section 5, good initialization improves convergence in small-sample regimes, though all algorithms eventually stabilize near the true values as data accumulate.

Well-specified setting: Additional results The same hyperparameter choices were made for the algorithms as in the well-specified setting in Section 5. Namely, we use $\epsilon_t = t^{-1.5}$ to encourage

Table 6: True and initial parameter values used in simulation studies.

Model	True Parameters	Initial Values
Mitscherlich	$A = 120, b = 0.015, d = 80$	$A = 100, b = 0.01, d = 75$
Quadratic Plateau	$a = 80, b = 1.2, c = -0.003, x_0 = 180$	$a = 75, b = 1.0, c = -0.002, x_0 = 160$
Michaelis–Menten	$a = 150, b = 100, d = 60$	$a = 120, b = 80, d = 50$
Logistic	$A = 120, B = 0.05, C = 125, d = 70$	$A = 100, B = 0.03, C = 100, d = 65$

exploitation sooner than later, $\alpha = 1$ for UCB and $\alpha_1 = 2, \alpha_2 = 640$ for Vi0lin. Each algorithm was run 10 times and $T = 30$ to emulate data-limited setting.

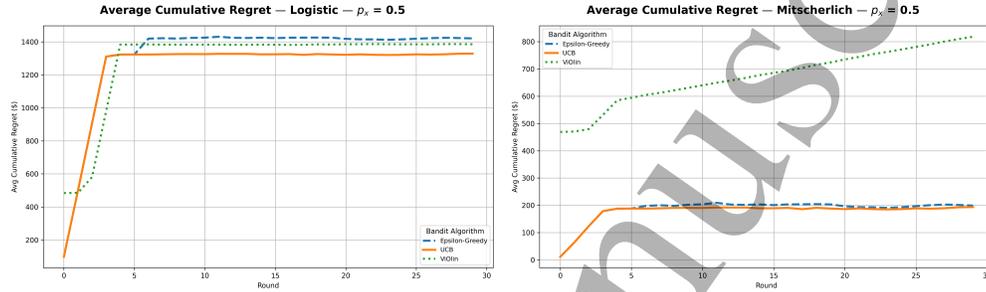


Figure 8: **Additional well-specified results** ($p_x = \$0.5/\text{lb N}$). Average cumulative profit regret (\$/ac) over $T = 30$ rounds for the logistic (left) and Mitscherlich (right) response models, comparing model-based ϵ -greedy ($\epsilon_t = t^{-1.5}$), nonlinear-UCB ($\alpha = 1$), and Vi0lin ($\alpha_1 = 2, \alpha_2 = 640$). Curves are averaged over 10 independent simulation runs.

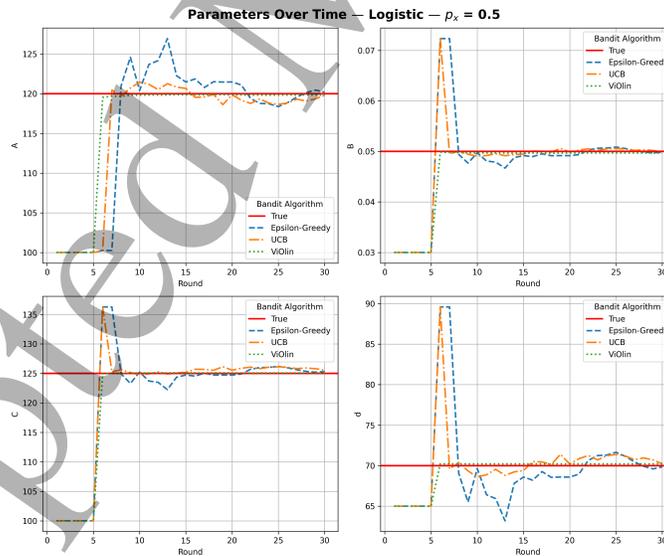


Figure 9: **Parameter-learning trajectories** (well-specified logistic model; $p_x = \$0.5/\text{lb N}$). Estimated logistic parameters over $T = 30$ rounds for model-based ϵ -greedy ($\epsilon_t = t^{-1.5}$), nonlinear-UCB ($\alpha = 1$), and Vi0lin ($\alpha_1 = 2, \alpha_2 = 640$). Red horizontal lines denote the true parameter values; one representative run is shown.

It can be noted from Figure 10 that while the arm choices remain mostly similar in the Logistic model to that of Quadratic plateau model, for the Mitscherlich model the Vi0lin algorithm seems

to choose $x = 250 \ln N/\text{ac}$ more than other choices, which seems also reflect in the worsened regret for Vi01in in Figure 8. We hypothesize that this is because, as the yield curve quickly saturates, the local curvature used for exploration in Vi01in becomes very small, causing the algorithm to underestimate uncertainty and prematurely stop exploring. Also, we notice greater variability in the estimation of parameters especially in the Logistic model as can be seen in Figure 9.

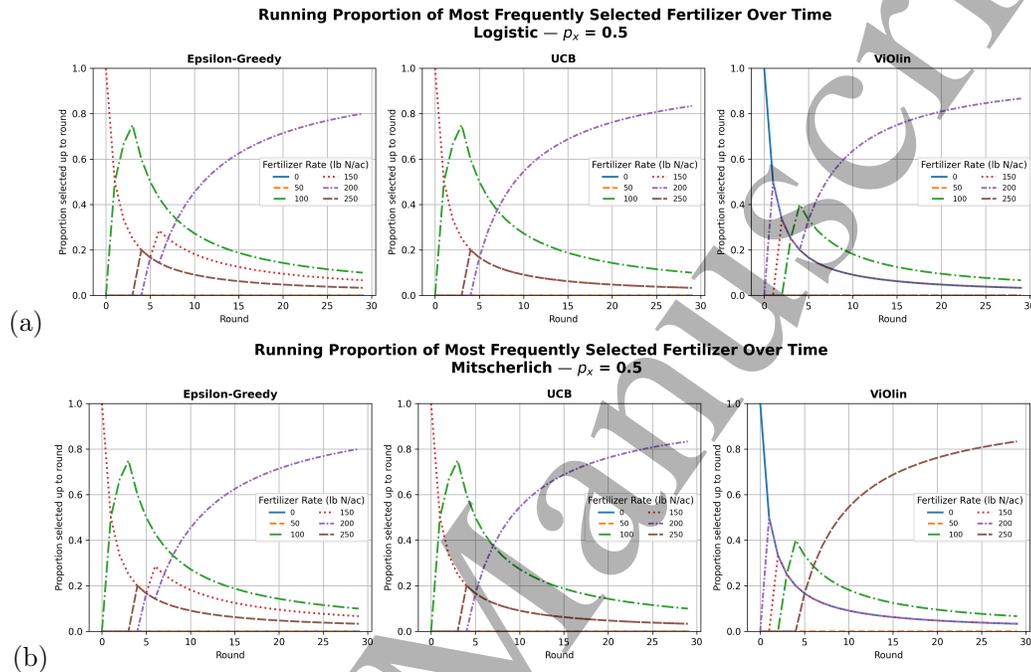


Figure 10: **Fertilizer-rate choice dynamics in the well-specified setting** ($p_x = \$0.5/\text{lb N}$). Running selection proportions over $T = 30$ rounds for the most frequently selected nitrogen rates (lb N/ac) under ϵ -greedy, nonlinear-UCB, and Vi01in. (a) Logistic response; (b) Mitscherlich response.

Misspecified setting: Additional results To complement the main-text regret comparisons (Figure 6), we provide additional diagnostics illustrating learning dynamics when the fitted model is misspecified relative to the true data-generating process. Figure 14 reports the trajectories of the fitted quadratic-plateau parameters over time when data are generated from a Mitscherlich response. Although the fitted functional form is imperfect, parameter estimates stabilize toward pseudo-true values that approximate the underlying exponential saturation curve on the discrete nitrogen grid. Figure 13 displays the evolution of fertilizer-rate selection frequencies across rounds. Consistent with the regret results, nonlinear model-based algorithms concentrate their selections on economically favorable regions of the action space after an initial exploration phase. These diagnostics illustrate that even under functional misspecification, shape-compatible parametric models can guide the bandit search toward profitable nitrogen rates in data-limited regimes.

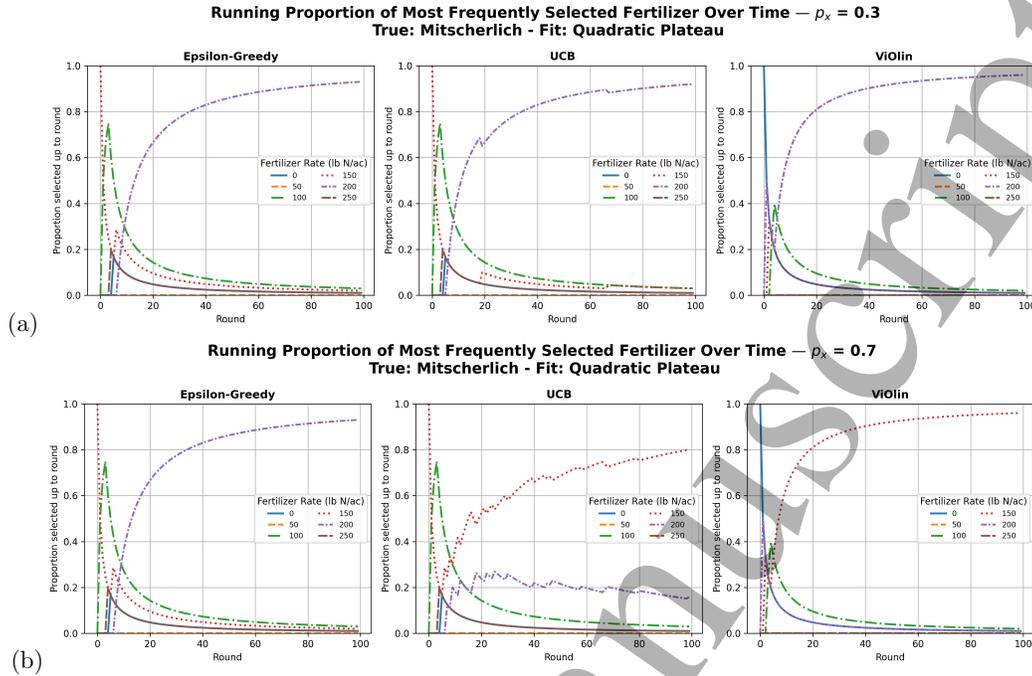


Figure 11: **Fertilizer-rate choice dynamics under misspecification (true Mitscherlich; fitted quadratic-plateau).** Running selection proportions for each nitrogen rate in $\mathcal{X} = \{0, 50, \dots, 250\}$ lb N/ac over $T = 100$ rounds (averaged over 10 replicates) for model-based ϵ -greedy ($\epsilon_t = t^{-1.5}$), nonlinear-UCB ($\alpha = 1$), and ViOLin ($\alpha_1 = 2, \alpha_2 = 640$). Panels compare two fertilizer-price regimes: (a) $p_x = \$0.3/\text{lb N}$ and (b) $p_x = \$0.7/\text{lb N}$. Higher fertilizer cost shifts the learned policies toward lower nitrogen rates.

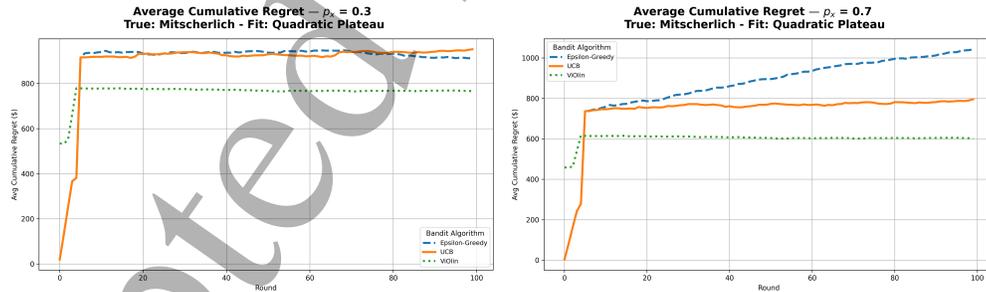


Figure 12: **Profit cumulative regret under model misspecification (true Mitscherlich; fitted quadratic-plateau).** Average cumulative regret (in $\$/\text{ac}$) over $T = 100$ rounds for model-based ϵ -greedy ($\epsilon_t = t^{-1.5}$), nonlinear-UCB ($\alpha = 1$), and ViOLin ($\alpha_1 = 2, \alpha_2 = 640$), averaged over 10 replicates. Panels compare two fertilizer-price regimes: left $p_x = \$0.3/\text{lb N}$ and right $p_x = \$0.7/\text{lb N}$.

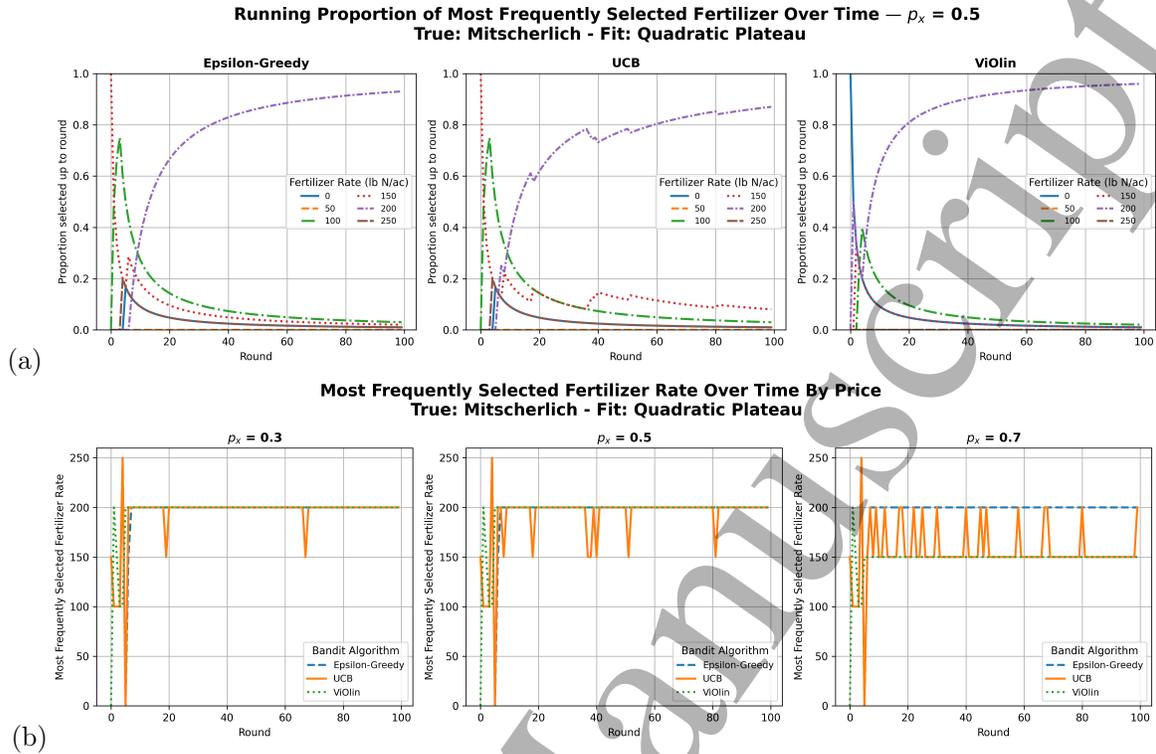


Figure 13: **Fertilizer-rate choice dynamics under misspecification (true Mitscherlich; fitted quadratic-plateau)**. Actions are chosen from $\mathcal{X} = \{0, 50, \dots, 250\}$ lb N/ac over $T = 100$ rounds. (a) For $p_x = \$0.5/\text{lb N}$, running selection proportions for each nitrogen rate are shown for ϵ -greedy, nonlinear-UCB, and Violin (averaged over 10 replicates). (b) The most frequently selected nitrogen rate is shown for $p_x \in \{0.3, 0.5, 0.7\}$ \$/lb N, highlighting how fertilizer price affects the learned policy when the fitted model is imperfect.

For the misspecified setting, for the same combination with the true model being Mitscherlich and fitted model being Quadratic plateau, we illustrate the effect of changing the level of fertilizer price from \$0.3 to \$0.7 per unit lb N. The same hyperparameter choices are made as in the misspecified setting of Section 5.

Figure 11 (a) corresponds to $p_X = 0.3$ while Figure 11 (b) corresponds to $p_X = 0.7$. Note that, both UCB and Violin capture that as the price increases the most frequently chosen arm over time becomes $x = 150$ (brown dotted line) instead of $x = 200$ (purple dashed dotted line). This sort of adaptive learning is a robustness check on how effective the decision-making is towards the goal of profit maximization. However, ϵ -greedy failed to adapt to this change, perhaps due to a fast decaying ϵ_t which made it get stuck on a sub-optimal arm in the initial rounds. This is also reflected in the poor regret performance by ϵ -greedy in Figure 12.

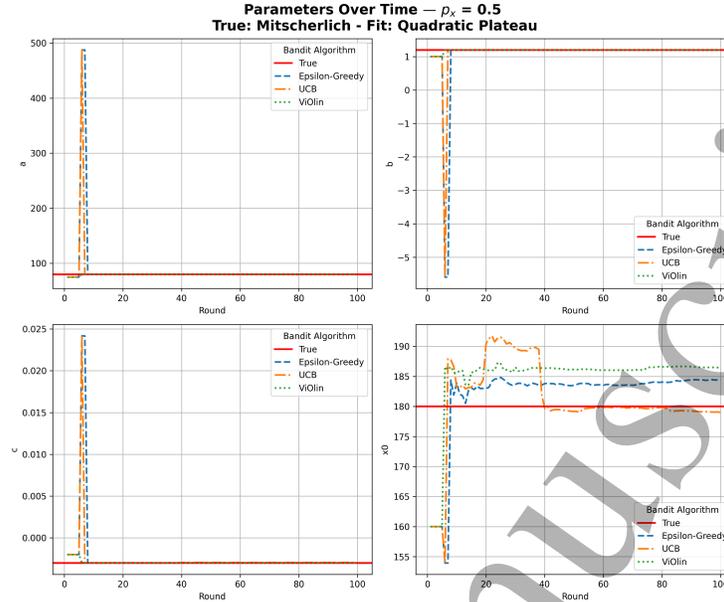


Figure 14: **Estimated parameter trajectories under misspecification (true Mitscherlich; fitted quadratic-plateau).** Shown for $p_x = \$0.5/\text{lb N}$ over $T = 100$ rounds. Panels track fitted quadratic-plateau parameters (a, b, c, x_0) for ϵ -greedy, nonlinear-UCB, and ViOlin. The red reference line denotes an *oracle (pseudo-true)* parameter value obtained by fitting the quadratic-plateau form to the noiseless Mitscherlich mean response on the same nitrogen grid \mathcal{X} (used only as a benchmark).

F Data description

We complement the simulation study with an offline evaluation on a multi-site nitrogen-rate trial dataset for corn production in the U.S. Midwest (Ransom et al., 2021). Our analysis uses the processed file `expanded_data.csv`, which aggregates all available trials across sites and states over 2014–2016. We also use restricted subsets of this file for illustrative case studies, e.g., Urbana, IL.

Experimental layout and key variables. The underlying field experiments follow a randomized complete block design (RCBD) within each site-year, with four spatial blocks ($\text{Block} \in \{1, 2, 3, 4\}$) that are geographically close. Each plot receives a planting nitrogen rate `Plant_N` (lb N/ac) taking values in $\{0, 40, 80, 120, 160, 200, 240, 280\}$, and the response is recorded as yield `Yield_Bu` (bu/ac). The dataset also includes an agronomically motivated baseline covariate `ExpectYield1`, representing the expected yield for that site based on prior yield history and growing conditions, and a site-level productivity label `Site_Prod` indicating relative productivity (high vs. low) within each state. Table 7 lists the variables retained in `expanded_data.csv` for the offline replay analysis.

Table 7: Column names in the processed field-trial dataset used for offline replay.

Column	Description (brief)
Trial#	Trial identifier
Year	Growing season year
State	U.S. state code
Site	Site/location name
Site_Prod	Site productivity class (e.g., low/high)
Block	Block index within site-year (RCBD block)
Plant_N	Nitrogen application rate (lb N/ac)
Yield_Bu	Observed corn yield (bu/ac)
ExpectYield1	Expected yield covariate (as provided in source data)

Low-productivity pooled subset (used in Section 6). To reduce heterogeneity and keep the evaluation aligned with a non-contextual bandit model, our pooled analysis restricts attention to site-years labeled `Site_Prod = low`. Table 8 summarizes the resulting subset by state. Here `#sites` counts unique (`State`, `Site`) pairs, `#rounds` counts unique (`State`, `Site`, `Year`, `Block`) decision rounds, and `#rows` is the number of plot-level observations.

Table 8: Summary of the pooled low-productivity real-data subset (2014–2016). The pooled subset totals 7 states, 11 sites, 56 rounds, and 442 rows.

State	# Sites	# Rounds	# Rows
IL	2	12	93
MN	2	12	94
IA	2	8	64
IN	1	8	64
WI	2	8	63
MO	1	4	32
NE	1	4	32
Total	11	56	442

Price series for the profit objective. For the profit objective $\Pi(x) = p_y Y(x) - p_x x$ used in the real-data replay, Table 9 reports the year-specific corn price p_y and nitrogen cost p_x . Corn price p_y is the U.S. annual average price received (\$/bu) from USDA–NASS Crop Values (2014–2016). Urea prices are December Midwest retail prices (\$/ton); nitrogen cost is computed as $p_x = (\$/\text{ton}) / (2000 \times 0.46)$ using urea’s 46% N analysis.

Table 9: Price series used for profit in the real-data offline replay (urea).

Year	Corn price p_y (\$/bu)	Urea price (\$/ton)	Nitrogen cost p_x (\$/lb N)
2014	3.70	485	0.527
2015	3.61	431	0.468
2016	3.40	343	0.373