Spatial Spline for Variable Rate Nitrogen Using On-Farm Experiment Data

W. Hence Duncan & B. Wade Brorsen, Oklahoma State University,
hence.duncan@okstate.edu

Abstract

The Data-Intensive Farm Management (DIFM) project at the University of Illinois enables producers to conduct site-specific nitrogen experiments, generating spatially structured on-farm data. Current DIFM-supported machine learning models often overfit, yielding widely varying variable rate nitrogen (VRN) recommendations. This research seeks to develop a fast, accurate, and commercially viable VRN algorithm using DIFM data. To accomplish this goal, the study develops a spatially varying parameter model that meets the following criteria: (i) the functional form is a plateau model with the plateau varying across the field, (ii) computational speed is sufficient for the model to be used on large datasets, and (iii) the model can incorporate prior information from past research. The method uses a semiparametric spatial spline approach, estimating a continuous response surface rather than estimating a different parameter at each site in the field. Estimation is implemented with restricted maximum likelihood (REML) in the mgcv package, which provides an empirical-Bayes approximation to posterior inference at a fraction of the computational cost of full Bayesian methods such as MCMC or INLA. Monte Carlo simulations evaluate accuracy, speed, and economic value. Across 1,000 simulated fields spanning a wide range of spatial heterogeneity, the spatial VRN approach recovered 77% of the economic value that uniform application loses relative to perfect knowledge of the field. The mean gain over uniform application was $40.88 per hectare, and the full simulation completed in approximately 15 minutes on a standard workstation. The approach offers a structured, computationally efficient alternative to the machine learning models currently used in DIFM, supporting stable and interpretable variable rate nitrogen recommendations.

Introduction

Precision agriculture technologies have generated unprecedented amounts of farm-level data, creating new opportunities for site-specific management. Nitrogen, one of the most expensive and environmentally significant inputs in crop production, remains a central focus of variable rate management. Field trials from the Data-Intensive Farm Management (DIFM) project have shown that nitrogen response varies significantly across space and time, suggesting potential gains from variable rate nitrogen (VRN) application (Bullock et al., 2019). Translating experimental data into profitable VRN recommendations remains challenging (Kitchen et al., 2020; Scharf et al., 2017). While promising, the current machine learning approaches often struggle with limited temporal data and produce unrealistic recommendations that vary excessively across fields. This study proposes and tests a spatio-temporal modeling framework that uses spatial splines to estimate the surface of a quadratic–quadratic–plateau nitrogen response function. The goal is to determine whether this approach can be a foundation for generating site-specific variable rate nitrogen recommendations.

Agricultural economists have long recognized that crop response functions exhibit spatial structure and that pooling information across space can improve estimation accuracy (Hurley et al., 2004; Anselin et al., 2004a). Early spatial approaches often relied on pre-defined management zones or spatial dummy variables, which imposed artificial boundaries and failed to capture gradual transitions in yield response (Miao et al., 2006; Schabenberger and Gotway, 2005). These methods lacked predictive power and were poorly suited for modeling continuous spatial variability. Geographically Weighted Regression (GWR) advanced the field by estimating local coefficients (Fotheringham et al., 2002); however, GWR results are highly sensitive to bandwidth selection and lack a unified framework for statistical inference, making hypothesis testing and model comparison difficult (Wheeler, 2014; Lambert and Cho, 2022). GWR also suffers from edge effects and assumes dense spatial sampling, often impractical in agronomic trials. Additionally, GWR cannot easily accommodate missing data and becomes infeasible when spatial coordinates change across years, which is common in on-farm experimentation (Patterson, 2024).

Field trials suggest that site-specific nitrogen strategies may reduce losses and improve efficiency, though results can vary considerably across locations (Bullock et al., 2019; Robertson and Vitousek, 2009). Despite these benefits, VRN adoption remains limited partly due to the lack of scalable, efficient decision tools (Schimmelpfennig, 2016).

Although site-specific nitrogen modeling has been explored for decades, many attempts to generate profitable VRN recommendations have fallen short. Studies have repeatedly found that variable rate application provides limited economic returns in fields with low spatial variability or noisy yield response data (Boyer et al., 2011; Stefanini et al., 2015; Larson et al., 2020). These findings suggest that profitability depends on model sophistication and whether yield response exhibits clear and consistent spatial patterns, which are often lacking in practice.

Bayesian spatial models such as kriging and spatially varying coefficient models have gained popularity for their ability to incorporate prior information and uncertainty. However, they often suffer from practical constraints. Park et al. (2024) and Patterson (2024) had to limit estimation to a subset of locations or a single year due to the computational burden of Bayesian Kriging. Even with advanced samplers like Hamiltonian Monte Carlo (HMC), scalability remains a barrier to commercial application (Poursina and Brorsen, 2023a). These challenges highlight a common issue in existing literature, statistically appealing models often fail to produce usable recommendations in large-scale, real-world settings.

Machine learning (ML) methods have been promoted as fast, flexible tools for VRN, but have drawbacks. Random forests and neural networks often overfit in small samples, lack agronomic structure, and do not incorporate economically meaningful constraints (Maimaitijiang et al., 2021; Wang et al., 2024). Their reliance on high resolution, multi-source data also limits practicality, as such data are expensive and time-consuming to collect on a per-field basis. Consequently, these models tend to work best in research settings rather than operational ones.

Zhang, Brorsen, and Duncan (2026) introduced a multidegree spline plateau model that is linear in its parameters, enabling efficient estimation with Integrated Nested Laplace Approximation (R-INLA). Building on this foundation, Duncan, Brorsen, and Zhang (2026) extended the model to allow the plateau coefficient to vary spatially across the field. Their results demonstrated that nonlinear nitrogen response functions could be recovered under simulated conditions while reducing computation time compared to traditional Markov Chain Monte Carlo (MCMC) methods, representing a major step forward in addressing the scalability issues that have hindered the adoption of Bayesian VRN models. However, because that model estimated separate parameters at each location, it required many coefficients, and the estimates became noisy with only a few years of data, revealing the need for additional restrictions to stabilize inference. The current study advances this line of work by modeling a continuous nitrogen response surface using spatial splines, reducing model complexity while preserving spatial flexibility.

This study models nitrogen response as a continuous surface that varies smoothly across space rather than estimating separate equations at each location. The model relies on spatial basis functions, which are predefined mathematical curves used to represent variation across a field. These functions are combined using weights, allowing the model to approximate spatial patterns with fewer parameters than models that assign unique coefficients to every point. This more compact structure improves computational efficiency and helps prevent overfitting. In addition, smoothing the nitrogen response surface introduces structure that stabilizes estimation, particularly when only a few years of data are available. These features make the model more scalable, stable, and better suited for site-specific nitrogen recommendations based upon on-farm experimentation. Because yield response often changes gradually across a field, spatial smoothing can capture real agronomic variation more accurately than models that impose rigid zones or overfit with high-resolution coefficients. This structure improves model interpretability and stability, making it more suitable for field-scale recommendations.

Research has shown that crop yield often varies smoothly over space, supporting the need for flexible spatial methods (Anselin et al., 2004b). While the use of spline-based spatial surfaces in agricultural modeling is still emerging, related work provides support for this approach. Finley (2011) shows that spatially varying coefficient models using spline-based structures can outperform traditional methods when spatial dependence is complex or nonstationary. Wen (2023) applies spatial smoothing in crop modeling, underscoring the broader value of continuous spatial methods even when splines are not explicitly used. These studies collectively motivate using spline-based surfaces to model spatial variation in nitrogen response.

To evaluate the proposed spatial spline model's performance, we conduct a Monte Carlo simulation using synthetic yield data generated from known spatial nitrogen response functions. The model is then estimated for each simulated dataset to assess how well it recovers the true underlying yield surface. We compare predicted nitrogen recommendations to true optima to evaluate accuracy and economic performance. This two-part analysis provides evidence on whether spatial spline models can generate stable, interpretable, and site-specific nitrogen recommendations suitable for commercial implementation.

Conceptual Framework

This study aims to calculate VRN recommendations by modeling spatially varying nitrogen response and identifying profit-maximizing nitrogen rates across a field. The underlying economic objective is to maximize expected profit at each field location i, where profit is defined as:

(1) $$\max_{\mathbf{N}} \sum_i E[\pi(\mathbf{N}_i, X_i)]$$

$N_i \in (0, 30,$ 60, 90, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260)

where **N** is a vector of nitrogen application rates across locations. The expected profit is:

(2) $$E[\pi(\mathbf{N}_i, X_i)] = py(\mathbf{N_i}, X_i) - cn_i - f$$

where E$[\pi(\mathbf{N}_i, X_i)]$ represents the expected profit as a function of nitrogen rate **N** and X_i is characteristic of location i, P price of corn, Y yield of corn, which is a function of nitrogen and location, C cost of nitrogen, and F is fixed cost.

The optimal nitrogen rate N_i is defined as the solution to the first-order condition:

(3)

$$\frac{\partial Y(N_i, X_i)}{\partial N} P = C$$

where the marginal revenue equals the marginal cost of nitrogen. Estimating a continuous nitrogen response surface across the field enables profit-maximizing nitrogen recommendations to be derived directly from the modeled surface without needing discrete zone definitions or local coefficient estimates.

Data

A stochastic plateau model was used to simulate corn yield responses to varying nitrogen rates over a 10-year period across a 30×30 spatial grid, resulting in 900 unique cells, each defined by distinct (x, y) coordinates. The simulation included twenty nitrogen rates (0, 34, 67, 101, 123, 134, 146, 157, 168, 179, 190, 202, 213, 224, 235, 246, 258, 269, 280, and 291 kg N/ha) to assess yield responses. Each nitrogen level is assigned to 45 grid sections each year. These rates are not considered optimal for maximizing yield or profit at any specific location. Instead, they reflect the standard structure used in DIFM trials, which prioritize equally spaced and equally replicated input levels to facilitate robust statistical analysis (Mullally et al., 2021). The yield response model calculates yield based on nitrogen rate, plateau level, and spatial and temporal effects. Parameter values are from Patterson (2024), who estimated this functional form from field trials.

The yield (Y_{it}) at location i, in year t, under nitrogen rate N_{it} was modeled as:

$$Y_{it} = \min(aN_{it} + b, P_{it}) + \varepsilon_t \tag{4}$$

where $a = 0.69$ is the marginal yield gain per nitrogen unit, $b = 98.18$ is the base intercept, P_{it} is the plateau yield for location i in year t. Plateau yields incorporate both spatially correlated base levels and year-specific temporal effects. The plateau yield P_{it} at location i in year t is modeled as:

$$P_{it} = \bar{P} + S_i + \epsilon_t \tag{5}$$

where $\bar{P}$ is average plateau, S_i is the spatial deviation at location i, and $\epsilon_t \sim N(0, \tau^2)$ representing yearly random effects shared across locations.

Yearly effects were simulated as random draws from a normal distribution with variance $\tau^2 = 235.95$. These yearly effects shift the plateau yield uniformly across all locations each year to account for changes in weather conditions and other yearly variations. Studies have reported standard deviations ranging from 630 to 1,700 kg/ha across different regions, supporting the plausibility of this parameter choice (Ray et al., 2016; Paulson et al., 2024).

The spatial component of P_{it} is modeled as a Gaussian random field. Spatial correlation is defined by the exponential covariance structure:

$$\text{Cov}(S_i, S_j) = sill * \exp\left(-\frac{d_{ij}}{range}\right) \tag{6}$$

where d_{ij} is the distance between locations i and j. The sill, range, and nugget parameters are drawn afresh for each simulation from |N(0, 16,000,000)|, |N(8, 2)|, and N(3,940, 3,940) respectively, matching the spatial-covariance distribution used in the companion R-INLA chapter (Duncan, Brorsen, and Zhang, 2026) so that the two estimators are evaluated on the identical distribution of plausible field structures. This produces a mean spatial standard deviation of roughly 2,750 kg/ha per field with substantial variation across simulations, spanning fields with very little spatial heterogeneity to fields with pronounced spatial structure. The resulting data for a single simulation can be found in Figures 1 and 2.

Procedure

This study adapts a quadratic-quadratic plateau model, an extension of Zhang, Brorsen, and Duncan's (2026) three-degree spline model, to incorporate spatial effects in nitrogen response estimation. Building on recent work that employed R-INLA to estimate spatially structured splines for large-scale variable-rate nitrogen recommendations (Duncan, Brorsen, and Zhang, 2026), this approach emphasizes greater model flexibility by allowing spatially varying coefficients within a penalized spline framework. The model is implemented using smooth interaction terms in R's mgcv package, enabling estimation of site-specific yield responses.

The model captures yield response to nitrogen application across three regions, separated by two nitrogen thresholds, N_1^* and N_2^*. The model is structured to be linear in its parameters, even though the underlying relationship between nitrogen and yield is nonlinear. This design allows efficient estimation using penalized likelihood techniques and ensures compatibility with spatial smoothing methods used in generalized additive modeling. Maintaining linearity in parameters also aids in model identifiability, reduces computational complexity, and facilitates the inclusion of spatially varying effects without requiring nonlinear optimization routines.

The first region, when nitrogen levels are low ($N_{it} \leq N_1^*$), follows a quadratic function. The middle nitrogen range ($N_1^* < N_{it} \leq N_2^*$) also follows a quadratic function, modeling diminishing returns as nitrogen levels increase. The last region ($N_{it} > N_1^*$) follows a plateau function, where additional nitrogen does not contribute to further yield gains. The model is

(7)

$$y_{it} = \begin{cases} \beta_0 + \beta_1 N_{it} + \beta_2 N_{it}^2 + \varepsilon_{it}, \text{if } N_{it} \leq N_1^* \\ \beta_3 + \beta_4 N_{it} + \beta_5 N_{it}^2 + \varepsilon_{it}, \text{if } N_1^* < N_{it} \leq N_2^* \\ \beta_6 + \varepsilon_{it}, \text{if } N_{it} > N_2^* \end{cases}$$

subject to restrictions imposing continuity at each knot:

$$\beta_0 + \beta_1 N_1^* + \beta_2 {N_1^*}^2 = \beta_3 + \beta_4 N_1^* + \beta_5 {N_1^*}^2$$

$$\beta_3 + \beta_4 N_2^* + \beta_5 {N_2^*}^2 = \beta_6,$$

differentiability at each knot:

$$\beta_1 + 2\beta_2 N_1^* = \beta_4 + 2\beta_5 N_1^*$$

$$\beta_4 + 2\beta_5 N_2^* = 0,$$

and a concavity constraint:

$$\beta_5 \leq 0, \text{if } N_1^* < N_{it} \leq N_2^*$$

where y_{it} is yield of observation i in year t, N_{it} is the nitrogen application rate, N_1^* and N_2^* are knots of the spline function that are assumed known at 146 kg N/ha and 247 kg N/ha, and ε_{it} is the error term. The concavity condition $\beta_5 \leq 0$ is not imposed as a hard constraint during estimation; in our simulations the condition held for the parameter ranges encountered.

Similar to Zhang, Brorsen, and Duncan's (2026) derivation, we express β_2, β_3, β_4, and β_5 in terms of β_0, β_1, and β_6, reducing the model to three free parameters: β_0, representing the yield intercept of the first quadratic, β_1 which is the linear coefficient of the second quadratic and β_6 is the yield plateau. The resulting constrained model from the derivation in the Appendix becomes

(8) $$y_{it} = \phi_0\beta_0 + \phi_1\beta_1 + \phi_6\beta_6 + \varepsilon_{it}$$

where ϕ_0, ϕ_1, and ϕ_6 are functions of N_{it}, N_1^*, and N_2^*, derived from the continuity and differentiability constraints. Equation (8) is linear in its parameters of β_0, β_1 , and β_6. ϕ_0, ϕ_1, and ϕ_6 are defined as:

$$\phi_0 = D_{1it}\left(1 + \frac{2N_{it}^2}{-2N_1^*N_2^*}\right) + D_{2it}\left(\frac{2{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*} + \frac{2N_{it}}{(N_1^* - N_2^*)} + \frac{2N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}\right)$$

$$\phi_1 = D_{1it}\left(N_{it} + \frac{N_1^* + N_2^*}{-2N_1^*N_2^*}N_{it}^2\right) + D_{2it}\left(\frac{N_1^*{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*} + \frac{N_1^*N_{it}}{(N_1^* - N_2^*)} + \frac{N_1^*N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}\right)$$

$$\phi_6 = D_{1it}\left(\frac{-2N_{it}^2}{-2N_1^*N_2^*}\right) + D_{2it}\left(\frac{-2N_1^*N_2^*}{-2(N_1^* - N_2^*)N_2^*} + \frac{-2N_{it}}{(N_1^* - N_2^*)} + \frac{-2N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}\right) + D_{3it}$$

To account for spatial heterogeneity in yield response, the model is estimated using penalized spline methods with the mgcv package version 1.9.1 in R version 4.4.1. In this formulation, only the plateau parameter β_6 varies spatially while β_0 and β_1 remain constant. Because the model is reparametrized to express yield as a linear function of β_0, β_1, and β_6, allowing β_6 to vary spatially allows β_2 through β_5 to vary spatially as well. In addition to spatial variation, the model incorporates temporal variation using independent and identically distributed Gaussian random effects (the "re" basis in mgcv), accounting for year-to-year yield fluctuations due to weather or other annual effects. The model is expressed as

(9)

$$y_{it} = \phi_0\beta_0 + \phi_1\beta_1 + \phi_6(s_i)\beta_6 + \varepsilon_{it}$$

where s_i, represents the spatial location of observation i, and year t. The function $\phi_6(s_i)$ introduces spatial variation in the yield plateau, modeled using thin-plate regression splines (Wood, 2003), while temporal variation is captured using independent and identically distributed Gaussian random effects, matching the data-generating process for year-to-year shocks (Wood, 2017). We apply a penalized smoothing approach (Eilers and Marx, 1996) to prevent the model from overfitting the data while still allowing enough flexibility to capture spatial patterns. The spatially varying plateau $\beta_6(s_i)$ is estimated as

(10) $$\beta_6(s_i) = f(s_i)$$

where $f(s_i)$ is a smooth function of spatial coordinates estimated using penalized splines. Penalization ensures a balance between flexibility and overfitting and is automatically selected using restricted maximum likelihood (REML) (Wood, 2017). The estimated smooth term follows

(11)

$$f(s_i) = \sum_{k=1}^{K} \gamma_k B_k(s_i)$$

where $B_k(s_i)$ are known basis functions evaluated at the spatial location s_i and γ_k are the associated coefficients. In thin-plate regression splines, each basis function contributes a localized component to the overall spatial surface. The coefficients γ_k determine the magnitude and shape of each component and are estimated by minimizing a penalized residual sum of squares. This penalization balances goodness-of-fit with smoothness, discouraging overly complex surfaces unless strongly supported by the data and reducing overfitting (Ruppert et al., 2003). The estimation procedure minimizes:

(12)

$$\hat{\gamma} = arg\min_{\gamma}\{\sum_{i,t}(y_{it} - \phi_0\beta_0 - \phi_1\beta_1 - \beta_6(s_i))^2 + \lambda\int(f"(s_i))^2 ds$$

where λ is a smoothing parameter that controls the complexity of the spatial surface and is also selected via REML.

This approach ensures that only the yield plateau varies spatially, while the nitrogen response in the first region and the quadratic response remain constant across locations. Uncertainty in the estimated spatial and temporal effects is quantified through approximate confidence intervals for the smooth terms, derived using a large-sample Bayesian approximation implemented via REML in the mgcv package. These intervals provide insight into the reliability of the estimated surfaces and are particularly useful for identifying locations with high spatial uncertainty. Effective degrees of freedom (EDF) are also reported for each smooth term. EDF reflects the complexity of the fitted spline: higher values indicate greater flexibility, while lower values reflect stronger penalization and smoother surfaces.

This formulation includes three parameters, β_0 the intercept representing expected yield with zero nitrogen; β_1, the marginal yield response in the low nitrogen levels, and $\beta_6(s_i)$, the plateau yield that varies across space to reflect local yield plateaus. While β_0 and β_1 are fixed across all locations and estimated directly, $\beta_6(s_i)$, is modeled as a smooth spatial function instead of a separate parameter for each location. This surface is expressed as a linear combination of known basis functions, with γ_k determining the shape and magnitude of spatial variation. These basis function weights are the parameters estimated in the spatial component of the model. By expressing $\beta_6(s_i)$ this way, the model captures spatial variability without overfitting to noise or relying on location specific parameters.

The site-specific optimal nitrogen rate N_i^* is determined by solving the first-order condition:

$$\frac{\partial Y(N_i, X_i)}{\partial N} p_c = p_n \tag{13}$$

where p_c is the price of corn per kilogram ($0.197) and p_n is the price of nitrogen per kilogram ($1.323). Although the model is reparametrized to estimate only three coefficients, calculating the optimal nitrogen rate requires using the original quadratic-quadratic-plateau yield function. This is necessary because the marginal product of nitrogen, which enters the first-order condition, depends on the underlying second-order terms β_2 and β_5, which are derived from the estimated parameters and the knot points. Substituting the piecewise yield function into the first-order condition leads to closed-form expressions for the optimal rate within each region. In the low-nitrogen region, the optimal rate is

$$N_i^* = \frac{\frac{p_n}{p_c - \beta_1}}{2\beta_2}, \text{if } N_i \leq N_1^*$$

and in the middle region:

$$N_i^* = \frac{\frac{p_n}{p_c - \beta_4}}{2\beta_5}, \text{if } N_i \leq N_2^*.$$

If $N_i > N_2^*$, the marginal revenue is zero and less than the marginal cost, so the optimal rate will never exceed N_2^*.

A Monte Carlo simulation evaluates how well the proposed method recovers the spatial yield plateau and recommends site-specific optimal nitrogen rates across multiple simulated datasets. Prediction accuracy is evaluated using root mean squared error (RMSE) between the actual and estimated yield plateau values and between the true and estimated site-specific optimal nitrogen rates. This measures how closely the model recovers the underlying spatial yield patterns. Because the true data-generating process is known, this approach allows for a direct comparison between model-based recommendations and decisions made with perfect information. The Value of Information (VOI) framework (Bullock and Lowenberg-DeBoer, 2007) is used to quantify the economic cost of imperfect information. VOI is calculated as the difference in expected profit between knowing the true yield response at each location and one that relies on the estimated spatial model to inform nitrogen recommendations.

Results

This study presents findings from 1,000 Monte Carlo simulations designed to evaluate a spline-based spatial-temporal model for estimating nitrogen response and generating site-specific recommendations. Each simulation generated 9,000 yield observations across a 30-by-30 spatial grid over ten years. The fitted model estimates a quadratic–quadratic–plateau nitrogen response function through three free parameters, allowing the plateau coefficient (β_6) to vary smoothly over space using penalized thin-plate regression splines. The slope parameters (β_0 and β_1) are re-estimated in each simulation but are held constant across spatial locations within a simulation. The full 1,000-simulation Monte Carlo completed in approximately 15 minutes on a standard workstation with the per-simulation fit distributed across parallel R workers, corresponding to roughly 10 seconds of single-thread CPU time per simulation. This timing confirms that the estimator meets the commercial-feasibility criterion stated in the introduction.

Across the 1,000 replications, the estimated yield intercept (β_0) averaged 6,140 kg/ha, and the marginal effect of nitrogen (β_1) averaged 34.9 kg yield/kg N. These results are displayed in Table 1. The effective degrees of freedom (EDF) associated with the spatial spline term for β_6 averaged 58.6, while the year random effect averaged 8.91 EDF, indicating that year-to-year shocks are well-identified rather than smoothed away.

Figure 3 displays the estimated plateau surfaces for the true yield plateau as shown in Figure 2. The estimated plateaus captured the general spatial pattern of yield potential across the field, accurately identifying both high- and low-yielding zones. Some localized smoothing was evident, particularly in regions with sharp yield gradients. The RMSE between estimated and true plateau values averaged 1,794 kg/ha with a standard deviation of 1,017 kg/ha across simulations (Table 2). True and estimated nitrogen response surfaces were also compared directly. The true mean plateau yield was 13,158 kg/ha, while the estimated mean plateau yield was 12,587 kg/ha, indicating that the model closely recovered the spatial yield potential. Across all simulations and locations, the estimator captured 69% of the true plateau standard deviation (estimated 2,491 vs. true 3,629 kg/ha), reflecting the penalized smoother's-controlled trade-off of variance for bias.

Optimal nitrogen rates were computed at each grid cell based on the estimated yield response and compared to the true profit-maximizing N rate derived from the simulation parameters. Figure 4 displays the true optimal nitrogen rate recommendations across the field, and Figure 5 shows the corresponding estimated optimal nitrogen rates generated by the spline-based model. Across 1,000 simulations, the RMSE between estimated and true optimal nitrogen rates averaged 22.6 kg N/ha with a standard deviation of 11.9 kg N/ha (Table 3). The maximum RMSE observed across simulations was 66.1 kg N/ha, reflecting occasional localized deviations in high-variance regions. The true mean optimal N rate was 206.1 kg N/ha, while the estimated mean optimal N rate was 209.9 kg N/ha, indicating minimal bias in site-specific economic recommendations. Comparing dispersion, the estimator captured 90% of the true spatial variation in optimal nitrogen (estimated SD 42.9 vs. true SD 47.5 kg N/ha). The spatial pattern of where to apply more nitrogen and where to apply less is recovered nearly completely despite the underrecovery of plateau variation. This decoupling reflects the geometry of the QQP profit function, which is relatively flat near its optimum, so moderate errors in the plateau estimate translate to much smaller errors in the recommended nitrogen rate.

The VOI is the difference in profit from using estimated nitrogen recommendations versus the true yield plateaus with corn price set to $0.197/kg and nitrogen to $1.323/kg. Using the estimation resulted in an average loss of $12.27/ha (Table 4). Across 1,000 simulations, the minimum VOI observed was $0.04/ha, and the maximum VOI observed was $115.53/ha. This range indicates that the model consistently produced near-optimal recommendations across the field, with minimal profit loss even in the worst simulations.

To benchmark the spatial VRN model against a non-spatial approach, a uniform nitrogen rate was computed for each simulation. The uniform rate was derived from the same fitted QQP model as the spatial VRN recommendations, with only the spatial differentiation of nitrogen varying between strategies, so the comparison isolates the value of spatial differentiation rather than the value of knowing the field perfectly. The profit loss from applying a uniform nitrogen rate rather than the true site-specific optima averaged $53.15/ha with a standard deviation of $72.60 (Table 5). The RMSE between the uniform nitrogen rate and the true site-specific optimal rates averaged 40.9 kg N/ha, substantially higher than the spatial VRN model's RMSE of 22.6 kg N/ha. As summarized in Table 6, the spatial VRN approach improved profit by an average of $40.88/ha and reduced nitrogen rate RMSE by 18.3 kg N/ha relative to the uniform rate strategy, demonstrating that the gains from spatially varying nitrogen application exceed those from applying a single field-average rate.

Overall, these results demonstrate that the spline-based spatiotemporal model accurately estimates nitrogen response surfaces, makes stable economic recommendations, and minimizes profit loss relative to the true site-specific optima, even under realistic spatial and temporal variability.

Conclusion

This study developed and evaluated a spline-based spatial-temporal model for predicting yield response to nitrogen and generating site-specific optimal nitrogen recommendations. By incorporating spatial variation into the yield plateau while maintaining a constrained quadratic-quadratic-plateau structure, the model achieved accuracy in estimating yield responses and economic optima. Across 1,000 Monte Carlo simulations, the model recovered yield plateaus with an RMSE of 1,794 kg/ha and predicted optimal nitrogen rates with an RMSE of 22.6 kg N/ha. The resulting VOI analysis indicated an average economic loss of $12.27/ha relative to decisions based on perfect information, confirming that the estimated recommendations closely approximated the true optimum.

Compared to a uniform nitrogen rate approach, the spatial VRN strategy reduced profit loss by $40.88/ha ($12.27 vs $53.15/ha relative to perfect information) while also improving nitrogen rate accuracy (RMSE of 22.6 vs 40.9 kg N/ha). Expressed as a share of the value at stake, the spatial VRN approach recovered 77% of the per-hectare profit that uniform application loses relative to perfect knowledge of the field. These results demonstrate that the economic and agronomic gains from spatially varying nitrogen application exceed those from field-average management, supporting the value of site-specific approaches even when model predictions are imperfect.

Several limitations should be acknowledged. The data-generating process uses a linear-response-plateau form, while the estimator fits a quadratic-quadratic-plateau spline; results therefore reflect performance under functional-form misspecification. Performance on data generated from other functional forms was not tested. The knot locations $N_1^* = 130$ and $N_2^* = 220$ were chosen based on Duncan, Brorsen, and Zhang (2026) sweep over knot placements; on real fields these would need to be selected from agronomic priors or estimated directly. The spatial covariance draws span a broad distribution of plausible fields, but specific agronomic settings such as highly erosive landscapes or fields with irrigation gradients may exhibit covariance structures outside this distribution.

The 1,000-simulation Monte Carlo evaluation completes in approximately 15 minutes on a standard workstation, corresponding to single-second per-field fit times. The estimator therefore meets the commercial-feasibility criterion stated in the introduction while delivering competitive accuracy. Overall, these findings demonstrate the potential of flexible spline-based methods for practical use in variable-rate nitrogen management. Although the results are promising, validation utilizing on-farm data is needed to assess performance under real-world conditions.

References

Anselin, L., R. Bongiovanni, and J. Lowenberg-DeBoer. (2004a). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. *American Journal of Agricultural Economics*, 86(3): 675–687.

Anselin, L., I. Syabri, and Y. Kho. (2004b). Spatial econometrics and spatial filtering. *Regional Science and Urban Economics*, 34(5): 491–510.

Boyer, C. N., B. C. English, D. M. Lambert, R. J. Roberts, J. A. Larson, and J. L. Martin. (2011). The profitability of precision nitrogen management for corn. *Agronomy Journal*, 103(2): 452–461.

Bullock, D. S., and J. Lowenberg-DeBoer. (2007). Using spatial analysis to study the values of variable rate technology and information. *Journal of Agricultural Economics*, 58(3): 517–535.

Bullock, D. S., M. Boerngen, H. Tao, B. Maxwell, J. D. Luck, L. Shiratsuchi, L. Puntel, and N. F. Martin. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, 111(6): 2736–2746.

Duncan, W. H., B. W. Brorsen, and Y. Zhang. (2026). Using R-INLA to estimate multi-degree spline plateau models for variable rate nitrogen. Unpublished manuscript.

Eilers, P. H. C., and B. D. Marx. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2): 89–121.

Finley, A. O. (2011). Comparing spatially varying coefficient models for analysis of ecological data with nonstationary and anisotropic covariance structure. *Environmental and Ecological Statistics*, 18(2): 135–157.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Hoboken, NJ: Wiley.

Hurley, T. M., K. Oishi, and G. L. Malzer. (2004). Estimating the potential value of variable rate nitrogen applications: A comparison of spatial econometric and geostatistical models. *Journal of Agricultural and Resource Economics*, 29(2): 175–199.

Kitchen, N. R., J. F. Shanahan, C. J. Ransom, C. J. Bandura, G. M. Bean, J. J. Camberato, P. R. Carter et al. (2020). A public-industry partnership for enhancing corn nitrogen research and datasets: Project description, methodology, and outcomes. *Agronomy Journal*, 112(1): 1–37.

Lambert, D. M., and W. Cho. (2022). Geographically weighted regression estimation of the linear response and plateau function. *Precision Agriculture*, 23(3): 377–399.

Larson, J. A., R. K. Roberts, B. C. English, S. L. Larkin, D. M. Lambert, and J. M. Reeves. (2020). Whole-farm financial impacts of variable rate nitrogen application for cotton production. *Precision Agriculture*, 21(2): 410–427.

Maimaitijiang, M., V. Sagan, S. Bhadra, C. Nguyen, T. C. Mockler, and N. Shakoor. (2021). A fully automated and fast approach for canopy cover estimation using super high-resolution remote sensing imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3: 219–226.

Miao, Y., D. J. Mulla, P. C. Robert, and J. A. Hernandez. (2006). Within-field variation in corn yield and grain quality responses to nitrogen fertilizer in Minnesota. *Agronomy Journal*, 98(3): 538–548.

Mullally, C., T. Cook, S. Dall'erba, and D. Bullock. (2021). Improving on-farm experimentation with spatial econometrics: Evidence from data-intensive farm management. *American Journal of Agricultural Economics*, 103(1): 215–239.

Park, J., S. Kim, and H. Lee. (2024). Bayesian kriging for optimal nitrogen recommendations in cornfields. *Environmental and Ecological Statistics.* (Advance online publication.)

Patterson, C. (2024). *Using Informative Bayesian Priors and On-Farm Experimentation to Predict Optimal Site-Specific Nitrogen Rates.* Master's thesis, Oklahoma State University.

Paulson, N., G. Schnitkey, and C. Zulauf. (2024). County yield variability and ARC-CO. *farmdoc daily*, Department of Agricultural and Consumer Economics, University of Illinois at Urbana–Champaign. Available at: https://farmdocdaily.illinois.edu.

Poursina, M., and B. W. Brorsen. (2023a). Estimation of a linear stochastic plateau model using Hamiltonian Monte Carlo: Application to site-specific nitrogen recommendations. *Environmental and Ecological Statistics*, 30: 467–485.

Poursina, M., and B. W. Brorsen. (2023b). Fully Bayesian economically optimal design for a spatially varying coefficient linear stochastic plateau model over multiple years. *Stochastic Environmental Research and Risk Assessment*, 38(3): 1089–1098.

Ray, D. K., J. S. Gerber, G. K. MacDonald, and P. C. West. (2016). Climate variation explains a third of global crop yield variability. *Scientific Reports*, 6: 33160.

Robertson, G. P., and P. M. Vitousek. (2009). Nitrogen in agriculture: Balancing the cost of an essential resource. *Annual Review of Environment and Resources*, 34: 97–125.

Ruppert, D., M. P. Wand, and R. J. Carroll. (2003). *Semiparametric Regression.* Cambridge, UK: Cambridge University Press.

Schabenberger, O., and C. A. Gotway. (2005). *Statistical Methods for Spatial Data Analysis.* Boca Raton, FL: Chapman & Hall/CRC.

Scharf, P. C., N. R. Kitchen, K. A. Sudduth, and J. G. Davis. (2017). Spatially variable corn nitrogen management: Status and opportunities. *Agronomy Journal*, 109(6): 2203–2212.

Schimmelpfennig, D. (2016). *Farm Profits and Adoption of Precision Agriculture.* USDA-ERS Economic Research Report No. 217.

Stefanini, F. M., M. Panten, and S. Asseng. (2015). Spatially variable nitrogen fertilizer response functions in corn and their implications for site-specific fertilizer management. *Precision Agriculture*, 16(5): 494–510.

Wang, J., Y. Wang, G. Li, and Z. Qi. (2024). Integration of remote sensing and machine learning for precision agriculture: A comprehensive perspective on applications. *Agronomy*, 14(9): 1975.

Wen, K. (2023). A semiparametric spatio-temporal model of crop yield trend and its implication to insurance rating. *Agricultural Economics*, 54(5): 662–673.

Wheeler, D. (2014). Geographically weighted regression. In M. M. Fischer and A. Getis (eds.), *Handbook of Applied Spatial Analysis*, pp. 465–486. Berlin: Springer.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B*, 65(1): 95–114.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R.* 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Zhang, Y., B. W. Brorsen, and W. H. Duncan. (2026). Nitrogen response modeling with a multidegree spline plateau approach. *Journal of Agricultural and Resource Economics*, 51(2): 238–262. https://doi.org/10.22004/ag.econ.369143

Appendix

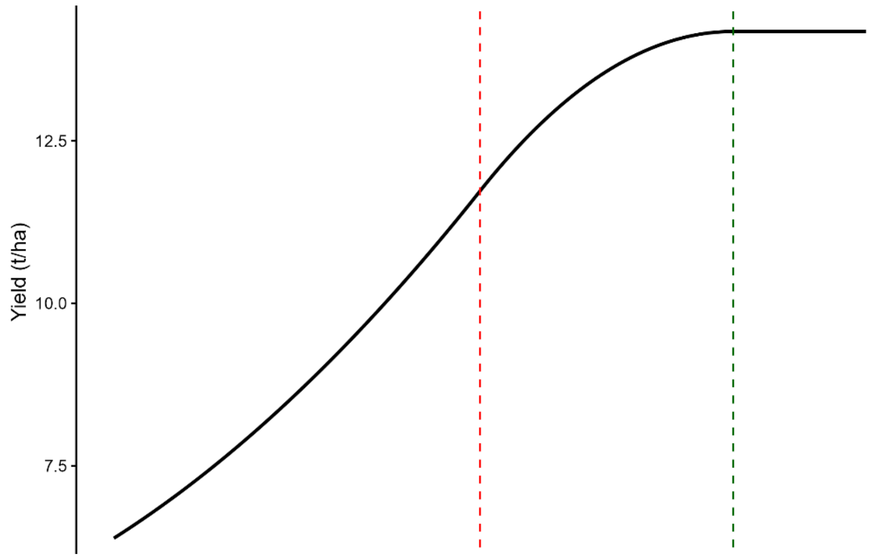

Figure 1. Average Estimated Yield Response Curve

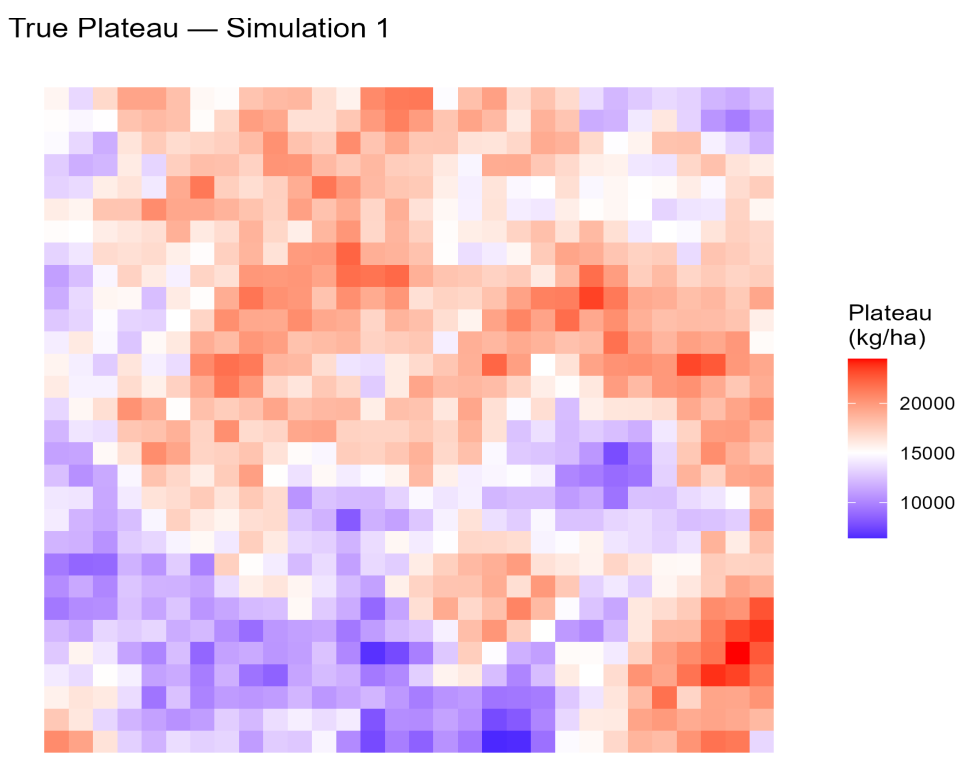

Figure 2. Example of Simulated Corn Plateau Yields

Table 1. Estimated Coefficients and Effective Degrees of Freedom (EDF) for the Spline Based Nitrogen Response Model Across 1,000 Simulations

Coefficient	Mean Estimate
β_0 (Intercept)	6,140
β_1 (Nitrogen Slope)	34.9
EDF: Spatial	58.6
EDF: Year (random effect)	8.91

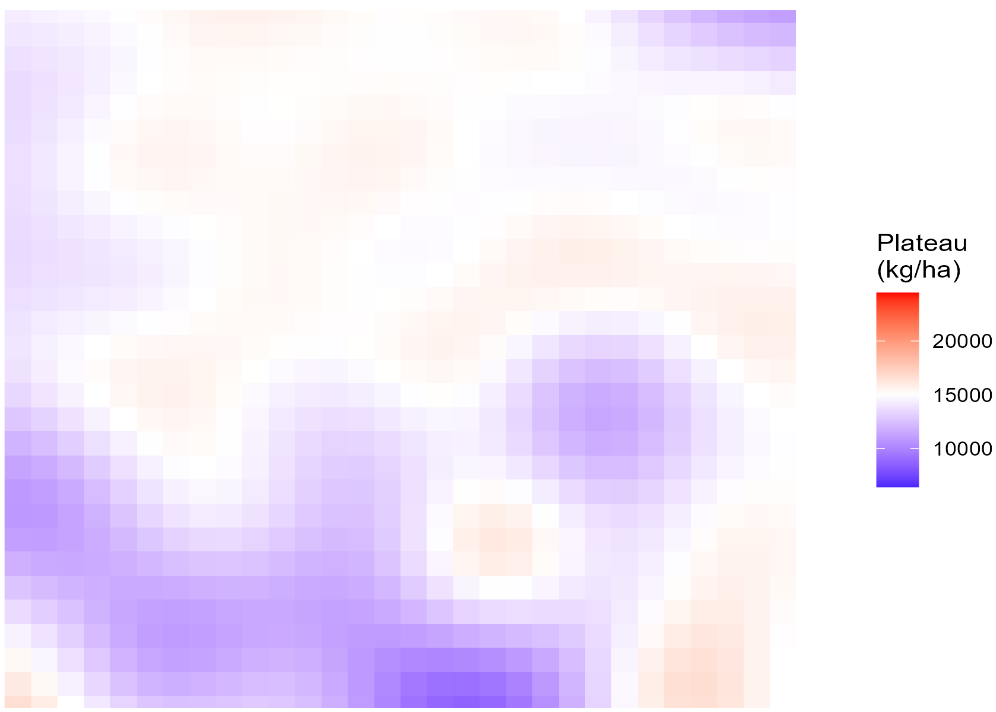

Figure 3. Example of Estimated Corn Plateau Yields

Table 2. Root Mean Squared Error (RMSE) Between True and Estimated Yield Plateaus Across 1,000 Simulations

Statistic	RMSE (kg/ha)
Mean RMSE	1,794
Standard Deviation	1,017
Minimum RMSE	237
Maximum RMSE	5,882

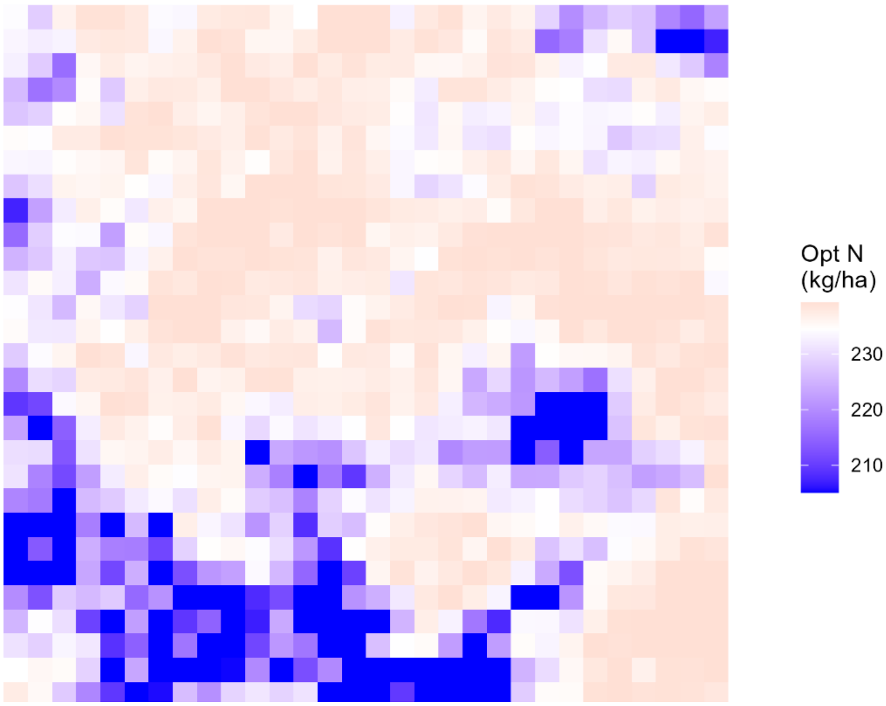

Figure 4. Example of True Optimal Nitrogen Rate Recommendations

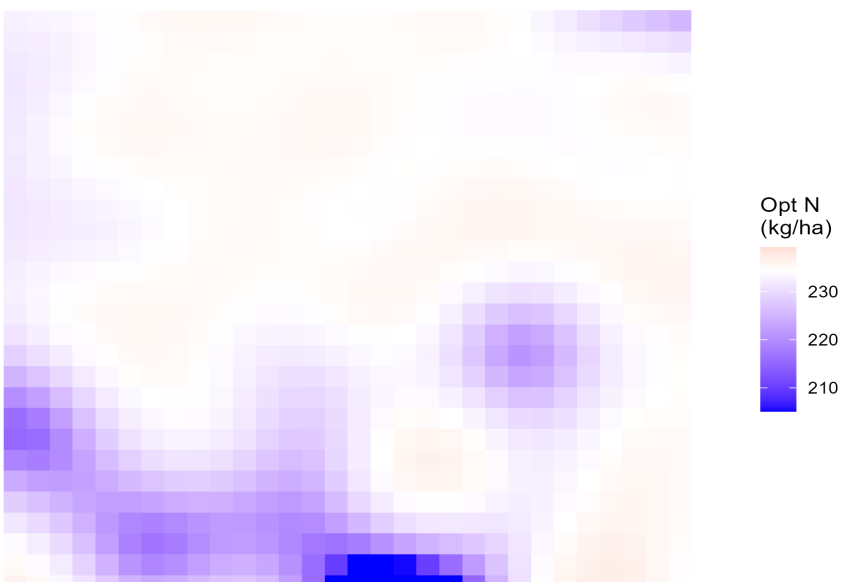

Figure 5. Example of Estimated Optimal Nitrogen Rate Recommendations

Table 3. Root Mean Squared Error (RMSE) Between True and Estimated Optimal Nitrogen Rates Across 1,000 Simulations

Statistic	RMSE (kg N/ha)
Mean RMSE	22.6
Standard Deviation	11.9
Minimum RMSE	1.0
Maximum RMSE	66.1

Table 4. Summary of Spatial Value of Information (VOI) Results Across 1,000 Simulations

Statistic	VOI ($/ha)
Mean	12.27
Standard Deviation	12.97
Minimum	0.04
Maximum	115.53

Table 5. Uniform Rate Performance Across 1,000 Simulations

Statistic	Value
Mean VOI ($/ha)	53.15
Standard Deviation	72.60
Minimum VOI ($/ha)	0.06
Maximum VOI ($/ha)	506.07
Mean RMSE (kg N/ha)	40.9

Table 6. Spatial VRN vs Uniform Rate

Strategy	Mean VOI ($/ha)	N Rate RMSE (kg N/ha)
Spatial VRN	12.27	22.6
Uniform Rate	53.15	40.9
Difference	40.88	18.3

Math Derivation

The three-degree spline plateau model can be represented in dummy variable form as

(A1) $$y_{it} = D_{1it}(\beta_0 + \beta_1 N_{it} + \beta_2 N_{it}^2) + D_{2it}(\beta_3 + \beta_4 N_{it} + \beta_5 N_{it}^2) + D_{3it}\beta_6 + \varepsilon_{it}$$

where y_{it} is the corn yield of observation i in year t, N_{it} is the nitrogen application rate of observation i in year t, D_{1it}, D_{2it}, and D_{3it} are dummy variables, D_{1it} equals 1 when the nitrogen rate N_{it} is less than or equal to N_1^*, D_{2it} equals 1 when the nitrogen rate N_{it} is greater than N_1^* and less than or equal to N_2^*, D_{3it} equals 1 when the nitrogen rate N_{it} is greater than N_2^*, and ε_{it} is a random error term that is normally distributed. The first part of the model (when $N_{it} \leq N_1^*$) is quadratic

(A2) $$y_{it} = \beta_0 + \beta_1 N_{it} + \beta_2 N_{it}^2 \varepsilon_{it}$$

The second part of the model (when $N_1^* < N_{it} \leq N_2^*$) is also quadratic

(A3) $$y_{it} = \beta_3 + \beta_4 N_{it} + \beta_5 N_{it}^2 + \varepsilon_{it}$$

And the third part (when $N_{it} > N_2^*$) is the plateau:

(A4) $$y_{it} = \beta_6 + \varepsilon_{it}$$

This model has five constraints, two constraints to make sure the model is continuous at the two knots N_1^* and N_2^*:

(A5) $$\beta_0 + \beta_1 N_1^* + \beta_2 {N_1^*}^2 = \beta_3 + \beta_4 N_1^* + \beta_5 {N_1^*}^2$$

(A6) $$\beta_3 + \beta_4 N_2^* + \beta_5 {N_2^*}^2 = \beta_6$$

two constraints to make sure the model is differentiable at the two knots:

(A7) $$\beta_1 + 2\beta_2 N_1^* = \beta_4 + 2\beta_5 N_1^*$$

(A8) $$\beta_4 + 2\beta_5 N_2^* = 0$$

Given the four constraints, β_2, β_3, β_4, and β_5 can be represented in terms of β_0, β_1 and β_6. From equation (A8), β_4 can be represented by β_5, as follows:

(A9)
$$\beta_4 = -2\beta_5 N_2^*$$

Substituting equation (A9) into equation (A6):

$$\beta_3 + (-2\beta_5 N_2^*)N_2^* + \beta_5 {N_2^*}^2 = \beta_6$$

(A10)
$$\beta_3 = \beta_5 {N_2^*}^2 + \beta_6$$

Substituting equation (A9) into equation (A7):

$$\beta_1 + 2\beta_2 N_1^* = -2\beta_5 N_2^* + 2\beta_5 N_1^*$$

$$2\beta_2 N_1^* = -2\beta_5 N_2^* + 2\beta_5 N_1^* - \beta_1$$

(A11)
$$\beta_2 = \frac{-2\beta_5 N_2^* + 2\beta_5 N_1^* - \beta_1}{2N_1^*}$$

Substituting equations (A9), (A10), and (A11) into equation (A11):

$$\beta_0 + \beta_1 N_1^* + \frac{-2\beta_5 N_2^* + 2\beta_5 N_1^* - \beta_1}{2N_1^*} {N_1^*}^2 = \beta_5 {N_2^*}^2 + \beta_6 - 2\beta_5 N_2^* N_1^* + \beta_5 {N_1^*}^2$$

$$\beta_0 + \beta_1 N_1^* + \beta_5 (N_1^* - N_2^*)N_1^* - \frac{\beta_1 N_1^*}{2} = \beta_5 (N_1^* - N_2^*)^2 + \beta_6$$

$$\beta_0 + \frac{\beta_1 N_1^*}{2} - \beta_6 = \beta_5 (N_1^* - N_2^*)^2 - \beta_5 (N_1^* - N_2^*)N_1^*$$

$$\beta_0 + \frac{\beta_1 N_1^*}{2} - \beta_6 = -\beta_5 (N_1^* - N_2^*)N_2^*$$

$$\beta_5 = \frac{\beta_0 + \frac{\beta_1 N_1^*}{2} - \beta_6}{-(N_1^* - N_2^*)N_2^*}$$

$$\beta_5 = \frac{2\beta_0 + \beta_1 N_1^* - 2\beta_6}{-2(N_1^* - N_2^*)N_2^*}$$

(A12) $$\beta_5 = \frac{2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2}{-2(N_1^* - N_2^*)N_2^*}\beta_6$$

Substituting equation (A12) into equation (A9):

$$\beta_4 = -2N_2^*(\frac{2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2}{-2(N_1^* - N_2^*)N_2^*}\beta_6)$$

(A13) $$\beta_4 = \frac{2}{(N_1^* - N_2^*)}\beta_0 + \frac{N_1^*}{(N_1^* - N_2^*)}\beta_1 + \frac{-2}{(N_1^* - N_2^*)}\beta_6$$

Substituting equation (A12) into equation (A10):

$$\beta_3 = (\frac{2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2}{-2(N_1^* - N_2^*)N_2^*}\beta_6){N_2^*}^2 + \beta_6$$

$$\beta_3 = \frac{2{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*}\beta_6 + \beta_6$$

(A14) $$\beta_3 = \frac{2{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*{N_2^*}^2}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2N_1^*N_2^*}{-2(N_1^* - N_2^*)N_2^*}\beta_6$$

Substituting equation (A12) into equation (A11):

$$\beta_2 = \frac{-2\beta_5 N_2^* + 2\beta_5 N_1^* - \beta_1}{2N_1^*}$$

$$\beta_2 = \frac{-\beta_1}{2N_1^*} + \frac{2N_1^* - 2N_2^*}{2N_1^*}\beta_5$$

$$\beta_2 = \frac{-\beta_1}{2N_1^*} + \frac{2(N_1^* - N_2^*)}{2N_1^*}(\frac{2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2}{-2(N_1^* - N_2^*)N_2^*}\beta_6)$$

$$\beta_2 = \frac{-\beta_1}{2N_1^*} + \frac{1}{2N_1^*}(\frac{2}{-N_2^*}\beta_0 + \frac{N_1^*}{-N_2^*}\beta_1 + \frac{-2}{-N_2^*}\beta_6)$$

(A15)
$$\beta_2 = \frac{2}{-2N_1^*N_2^*}\beta_0 + \frac{N_1^* + N_2^*}{-2N_1^*N_2^*}\beta_1 + \frac{-2}{-2N_1^*N_2^*}\beta_6$$

Therefore, the model can be written as:

(A16)
$$\begin{aligned} y_{it} = D_{1it}(\beta_0 + \beta_1 N_{it} + (\frac{2}{-2N_1^*N_2^*}\beta_0 + \frac{N_1^* + N_2^*}{-2N_1^*N_2^*}\beta_1 + \frac{-2}{-2N_1^*N_2^*}\beta_6)N_{it}^2) \\ + D_{2it}(\frac{2N_2^{*2}}{-2(N_1^* - N_2^*)N_2^*}\beta_0 + \frac{N_1^*N_2^{*2}}{-2(N_1^* - N_2^*)N_2^*}\beta_1 \\ + \frac{-2N_1^*N_2^*}{-2(N_1^* - N_2^*)N_2^*}\beta_6 + (\frac{2}{(N_1^* - N_2^*)}\beta_0 + \frac{N_1^*}{(N_1^* - N_2^*)}\beta_1 \\ + \frac{-2}{(N_1^* - N_2^*)}\beta_6)N_{it} + (\frac{2}{-2(N_1^* - N_2^*)N_2^*}\beta_0 \\ + \frac{N_1^*}{-2(N_1^* - N_2^*)N_2^*}\beta_1 + \frac{-2}{-2(N_1^* - N_2^*)N_2^*}\beta_6)N_{it}^2) + D_{3it}\beta_6 \\ + \varepsilon_{it} \end{aligned}$$

For estimation, the required function here is:

(A17)
$$y_{it} = \delta_0\beta_0 + \delta_1\beta_1 + \delta_6\beta_6 + \varepsilon_{it}$$

From equation (A16),

(A18)
$$\begin{aligned} \delta_0 = D_{1it}\left(1 + \frac{2N_{it}^2}{-2N_1^*N_2^*}\right) + D_{2it}(\frac{2N_2^{*2}}{-2(N_1^* - N_2^*)N_2^*} + \frac{2N_{it}}{(N_1^* - N_2^*)} \\ + \frac{2N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}) \end{aligned}$$

(A19)
$$\begin{aligned} \delta_1 = D_{1it}\left(N_{it} + \frac{N_1^* + N_2^*}{-2N_1^*N_2^*}N_{it}^2\right) + D_{2it}(\frac{N_1^*N_2^{*2}}{-2(N_1^* - N_2^*)N_2^*} + \frac{N_1^*N_{it}}{(N_1^* - N_2^*)} \\ + \frac{N_1^*N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}) \end{aligned}$$

$$\begin{aligned}\delta_6 = {} & D_{1it}\left(\frac{-2N_{it}^2}{-2N_1^* N_2^*}\right) \\ & + D_{2it}\left(\frac{-2N_1^* N_2^*}{-2(N_1^* - N_2^*)N_2^*} + \frac{-2N_{it}}{(N_1^* - N_2^*)} + \frac{-2N_{it}^2}{-2(N_1^* - N_2^*)N_2^*}\right) \\ & + D_{3it}\end{aligned} \tag{A20}$$