Optimizing Experimental Design for Determining Economic Nitrogen Levels: Insights on the use of Monte Carlo Simulations

Custódio Efraim Matavel[1]; Andreas Meyer-Aurich[1]; Hans-Peter Piepho[2]

[1] Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Max-Eyth-Allee 100, 14469 Potsdam, Germany

[2]Universität Hohenheim, Institute of Crop Science, Biostatistics Unit, 70599 Stuttgart, Germany

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

The determination of economic nitrogen levels is a pivotal element in the quest for sustainable agricultural practices. Designing experiments to accurately identify these levels, especially in contexts constrained by limited plot availability, poses a significant challenge. In response to these challenges, this study endeavors to demonstrate an approach to optimize the experimental design for identifying economic nitrogen levels, even under such constraints. We employed statistical techniques, including parametric bootstrapping and Monte Carlo simulations, to rigorously test various experimental designs. The primary criterion guiding our evaluation was to identify the experimental design that exhibited the lowest variance. This focus on minimizing variance is crucial, as it ensures greater reliability and reproducibility of the results, ultimately leading to more accurate determinations of economic nitrogen levels. Our methodology provided insightful results, demonstrating that certain experimental layouts significantly enhance the accuracy and efficiency of nitrogen level experiments. The superiority of these designs was evident in their ability to consistently produce results with lower variance compared to other tested layouts. This is particularly valuable in contexts where resources and land for extensive experimental setups are scarce. This research not only presents a framework for more informed decision-making in agricultural experimental design but also showcases the potential for wider applications of simulation-based methodologies in agronomic research. The methods we employed can be adapted for various studies, including those focusing on optimal nitrogen application and site-specific fertilizer utilization. The implications of our findings are substantial for the agricultural sector. By enabling more precise and efficient nitrogen management, our approach contributes to better crop yields, reduced environmental impact, and optimized resource utilization. In a broader context, this aligns with global goals for sustainability and food security. Our study thus offers a valuable contribution to the field of agricultural research, presenting a solution to a long-standing problem and paving the way for future advancements in sustainable farming practices.

Keywords.
EONR; Agricultural Experimentation; Monte Carlo Simulation; Optimal Fertilizer Application.

Introduction

Identifying the economic optimal nitrogen rate (EONR) is crucial for sustainable agriculture, aiming to maximize crop yields while minimizing environmental impact. Field experiments are instrumental in determining the precise nutrient application needed, helping stakeholders pinpoint the most cost-effective rates (Meyer-Aurich et al., 2010; Poursina et al., 2024). These trials are key to the customized management of agricultural inputs, enhancing farm-level efficiency and sustainability, and ultimately improving economic returns.

Nevertheless, the ideal configuration of these trials is a subject of ongoing debate, especially concerning economic factors such as the choice of treatment levels—referred to as 'design points'—their spatial distribution, and the required number of repetitions to ensure statistical robustness and cost-efficiency (Poursina & Brorsen, 2023; Poursina et al., 2024). To tackle these issues, researchers are increasingly employing Optimal Experimental Design (OED) techniques, which involve carefully selecting design points and their repetitions to enhance the efficiency and economic feasibility of experiments (Bachmaier, 2012; Büchse et al., 2022; Ng'ombe & Brorsen, 2022; Piepho et al., 2011; Poursina et al., 2023).

However, the advantages of OED depend heavily on the accuracy of the underlying model assumptions. Incorrect assumptions can lead to inefficient and economically suboptimal experiments (Alhorn et al., 2021). Therefore, refining model accuracy is essential, and experimental data is vital in this process. Models that are inaccurate may need adjustments, whether by adopting more complex or simpler structures, to improve their predictive capabilities. The objective is to design experiments that yield informative data, facilitating precise model evaluations (Tarantola, 2005).

For non-linear systems, OED has shown considerable benefits over traditional Design of Experiments (DoE). In these contexts, OED can more effectively reduce parameter variance, making it an essential tool in fields where precise parameter estimation is crucial (Franceschini & Macchietto, 2008). However, because OED relies on the Fisher Information Matrix (FIM), it can sometimes miss the effects of model non-linearities. Other methods like Monte Carlo simulations can address these gaps (Moles et al., 2003; Sin et al., 2009).

Monte Carlo simulations provide a comprehensive framework for managing the uncertainties and complexities of non-linear models. They offer a more precise method for quantifying parameter uncertainty compared to traditional linear approaches by considering the full range of variability in experimental data (López C et al., 2015; Moles et al., 2003). This method is particularly advantageous in agricultural research, where the dynamic and intricate nature of biological systems often defies simpler analytical techniques.

In this research, we aim to refine the experimental design for determining EONR using Monte Carlo simulations. By concentrating on variance reduction, we aim to enhance the reliability and reproducibility of the results. Our approach involves evaluating various experimental designs under constrained conditions and assessing their ability to produce low-variance outcomes. This emphasis on variance reduction is critical, as it directly influences the accuracy and efficiency of nitrogen application trials.

The findings from this research highlight that specific experimental configurations can greatly improve the precision of nitrogen management practices, even with limited resources. This has significant implications for the agricultural sector, potentially leading to increased crop yields, lower environmental impacts, and better resource utilization. Our methodology not only offers a framework for more informed decisions in agricultural experimental design but also demonstrates the broader applicability of simulation-based methods in agronomic research.

Methodology

Problem Formulation

In our context, optimal design involves selecting a design parameter ξ that maximizes the expected utility of an experiment. The model for the experiment is defined by the distribution $p(y|\beta,\xi)$, where y represents the observable data, and β denotes the unknown parameters. The prior distribution for β is $p(\beta)$ and the utility function $U(y,\beta,\xi)$ quantifies the benefit of the outcomes. Our primary goal is to minimize the variance of the estimated EONR.

The expected utility is expressed as:

$$E[U(\xi)] = \int U(y,\beta,\xi)p(y|\beta,\xi)p(\beta)dyd\beta \tag{1}$$

Reparameterizing Models for EONR Estimation

To incorporate EONR directly into our analysis, we reparameterized four commonly used yield response functions: quadratic, quadratic plateau, Mitscherlich, and linear plateau. This reparameterization ensures that EONR can be directly estimated within each model.

Mitscherlich Model:

$$f(N) = \beta_0 + (\beta_1 - \beta_0)e^{-\beta_2 N} \tag{2}$$

Reparameterized as:

$$f(N) = \beta_0 + \left(-\frac{r}{p\beta_2 e^{-\beta_2 EONR_M}}\right)e^{-\beta_2 N} \tag{3}$$

Quadratic Model:

$$f(N) = \beta_0 + \beta_1 N + \beta_2 N^2 \tag{4}$$

Reparameterized as:

$$f(N) = \beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_Q\right)N + \beta_2 N^2 \tag{5}$$

Quadratic Plateau Model:

$$f(N) = \begin{cases} \beta_0 + \beta_1 N + \beta_2 N^2, \ if\ N < N^* \\ \beta_0 + \beta_1 N^* + \beta_2 (N^*)^2, \ if\ N \geq N^* \end{cases} \tag{6}$$

Reparameterized as:

$$f(N) = \begin{cases} \beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_Q\right)N + \beta_2 N^2, \ if\ N < N^* \\ \beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_Q\right)N^* + \beta_2 (N^*)^2, \ if\ N \geq N^* \end{cases} \tag{7}$$

Linear Plateau Model:

$$f(N) = \begin{cases} \beta_0 + \beta_1 N, \; if \; N < N^* \\ \beta_0 + \beta_1 N^*, \; if \; N \geq N^* \end{cases} \tag{8}$$

Reparameterized as:

$$f(N) = \begin{cases} \beta_0 + \beta_1 N, \; if \; N < EONR_{LP} \\ \beta_0 + \beta_1 EONR_{LP}, \; if \; N \geq EONR_{LP} \end{cases} \tag{9}$$

Model Averaging

Given the uncertainty in identifying the most accurate yield response model, we adopted a model averaging approach. This technique synthesizes predictions from multiple models to produce a more robust and reliable estimate of yield.

The average model is expressed as:

$$f(N) = w_q\left(\beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_Q\right)N + \beta_2 N^2\right) + w_m\left(\beta_0 + \left(-\frac{r}{p\beta_2 e^{-\beta_2 EONR_M}}\right)e^{-\beta_2 N}\right) + w_{qp}\begin{pmatrix} \beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_{QP}\right)N + \beta_2 N^2, \; if \; N < N^* \\ \beta_0 + \left(\frac{r}{p} - 2\beta_2 EONR_{QP}\right)N^* + \beta_2 (N^*)^2, \; if \; N \geq N^* \end{pmatrix} + w_{lp}\begin{pmatrix} \beta_0 + \beta_1 N, \; if \; N < EONR_{LP} \\ \beta_0 + \beta_1 EONR_{LP}, \; if \; N \geq EONR_{LP} \end{pmatrix} \tag{10}$$

Simulation Strategies

Monte Carlo simulations

We generated synthetic data using Monte Carlo simulations. The parameters for these simulations were estimated from an analysis of real-world agricultural data obtained from an N application experiment conducted in a completely randomized block design at the Sieblerfeld test field (5 ha) in Upper Bavaria, Germany. This experiment provided a dataset from which we extracted valuable insights by fitting reparameterized functions. For a detailed exposition of the experimental setup, readers are referred to (Bachmaier & Gandorfer, 2009).

Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) methods were employed to generate samples from the synthetic data distributions. This approach is essential for accurately estimating the parameters and expected utility, even when direct sampling of the posterior distribution is challenging due to its complexity. The expected utility is approximated using:

$$\hat{E}[U(\xi)] = \frac{1}{N}\sum_{i=1}^{N} U(y, \beta_i, \xi) \tag{11}$$

Where β_i are samples from the posterior distribution obtained via MCMC.

Augmented Probability Models

Augmented Probability Models (APMs) are used to facilitate efficient sampling and estimation in complex Bayesian models. By introducing auxiliary variables or extending the parameter space, APMs help in managing the computational complexity and improving the convergence of MCMC methods. In our study, we defined an augmented joint distribution $p(\xi, y, \beta, \eta)$ where η represents the auxiliary variables.

1- Model Augmentation:

- Extend the parameter space to include additional latent variables η o capture unobserved variability in yield response.
- Define the augmented joint distribution $p(\xi, y, \beta, \eta)$ such that it simplifies the sampling process.

2- Sampling:
- Use MCMC to generate samples from the augmented joint distribution.
- This approach improves the efficiency and accuracy of sampling by accounting for the auxiliary variables.

3- Marginalization:
- Integrate out the auxiliary variables η to obtain the marginal distribution of the original parameters and design variables.
- This step ensures that the results align with the original problem formulation.

Simulated Annealing

To address the challenge of high-dimensional expected utility surfaces, we applied a simulated annealing approach. The high dimensionality in this context arises from the complex, multi-layered structure of the average model, which combines multiple yield response functions, each with its own set of parameters. Additionally, the experimental design includes multiple plots and nitrogen levels, further increasing the dimensionality. Simulated annealing iteratively adjusts the nitrogen levels to focus sampling on regions with higher expected utility, facilitating the identification of the optimal design. Here's how it works in the context of our study:

1- Initialization: Start with an initial experimental design, characterized by a set of nitrogen levels.
2- Control Variable (Temperature): Begin with a high value for the control variable T. This variable governs the probability of accepting suboptimal solutions during the optimization process, allowing the algorithm to explore a wide range of design configurations.
3- Design Adjustment: Generate new candidate designs by making small adjustments to the nitrogen levels in the current design.
4- Utility Evaluation: Compute the expected utility for each candidate design using the reparameterized and averaged yield response models. The utility function is designed to minimize the variance of the EONR estimates.
5- Acceptance Probability: Accept or reject the new design based on its utility. Even if a new design has a higher variance (lower utility), it might still be accepted to avoid getting trapped in local optima. The acceptance probability is given by:

$$P(accept) = \exp\left(\frac{-\Delta U}{T}\right) \tag{12}$$

where ΔU is the change in utility and T is the current value of the control variable.

6- Cooling Schedule: Gradually reduce the control variable T according to a predefined schedule, focusing the search on refining the best designs found.
7- Convergence Check: Continue the process until the control variable T is sufficiently low and the design parameters stabilize, indicating that the optimal or near-optimal design has been found.

Experimental Setup

We considered a limited number of plots, specifically 120 plots, to optimize our experimental setup. We generated designs with 4, 5, 6, 8, and 10 levels of nitrogen from a design space varying between 0 and 300 kg N/ha, creating a variety of experimental configurations to identify the optimal design.

Implementation

Our methodology was implemented using PyMC3, a Python library for Bayesian statistical modeling. The simulation steps were as follows:

1- Model Specification: Define the probabilistic model and utility function.

2- Sampling: Generate synthetic data samples using Monte Carlo simulations and use MCMC for sampling from the synthetic distributions.
3- Utility Calculation: Compute the utility for each sampled design using the reparameterized and averaged model.
4- Optimization: Apply simulated annealing to find the design that maximizes the expected utility while minimizing the variance of EONR.

Results and discussion

The results of our simulations highlight the effectiveness of selecting optimal nitrogen levels in reducing the variance of EONR. For the design with 4 nitrogen rates, the optimal levels were found to be 10, 175, 211, and 285 kg N/ha (Figure 1). These levels showed significantly lower variances compared to other levels within the design space. When increasing the number of nitrogen rates to 5, the levels 20, 70, 210, 220, and 275 kg N/ha also resulted in lower variances within the design space.

For the design featuring 6 nitrogen rates, the optimal levels were identified as 50, 60, 172, 187, 225, and 267 kg N/ha, achieving notably low variances. Similarly, the designs with 8 nitrogen rates (29, 68, 93, 136, 154, 159, 172, and 255 kg N/ha) and 10 nitrogen rates (39, 56, 62, 113, 163, 208, 216, 226, 234, and 243 kg N/ha) showed that the optimal levels maintained lower variances compared to the rest of the design space.

It is essential to highlight that the limits set for the design space, which in our case ranged from 0 to 300 kg N/ha, should be carefully chosen to reflect environmentally sound practices. These limits should not be arbitrary; they must be derived from an understanding of the ecological consequences of excessive nitrogen use, including nutrient runoff that can lead to eutrophication and the degradation of water quality (Sharpley et al., 1987). Ensuring that the experimental designs adhere to environmentally sustainable nitrogen levels is crucial for the broader application of these findings in practical farming scenarios.

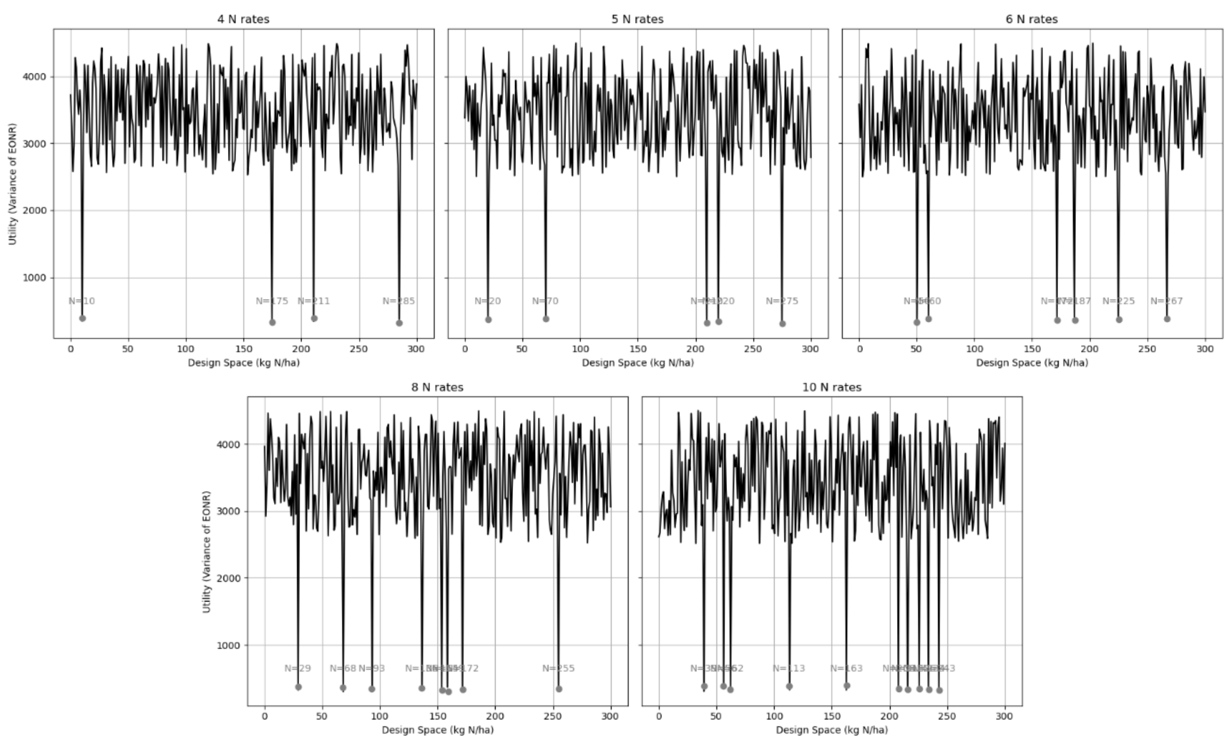

Fig 1. Utility (Variance of EONR) Across the Design Space with Optimal Nitrogen Levels.

The results of our study provide valuable insights for agricultural researchers and practitioners, particularly in contexts where resources and land for experimental setups are limited. By using this approach, it is possible to pinpoint the best designs that minimize the variance of EONR, thereby enhancing the accuracy and efficiency of nitrogen management practices. Our findings are consistent with the studies by Gilmour and Trinca (2012) and Pannell (1997), who emphasized the critical role of managing variance in agricultural experiments. The superior performance of the identified optimal levels in reducing variance aligns with global goals for sustainability and food security, as it enables more precise and efficient nitrogen management.

Our analysis demonstrates the critical role of selecting optimal nitrogen levels to minimize the variance of EONR, particularly under constraints of limited plot availability. While our results are based on a simulation approach, practical application is essential to fully understand the real-world applicability of these findings. This study offers a robust framework for improving the accuracy of nitrogen management practices in agriculture. By enabling more precise and efficient nitrogen management, our approach contributes to better crop yields, reduced environmental impact, and optimized resource utilization, aligning with global sustainability goals.

Conclusion

Determining economic nitrogen levels is crucial for sustainable agricultural practices, especially in contexts with limited plot availability. This study demonstrated an approach to optimize experimental designs for identifying these levels by employing statistical techniques, including parametric bootstrapping and Monte Carlo simulations. Our primary goal was to identify designs that minimized the variance of EONR estimates, ensuring greater reliability and reproducibility.

Our results showed that certain experimental layouts significantly enhance the accuracy and efficiency of nitrogen level experiments by consistently producing lower variances. This is particularly valuable in resource-constrained settings. The methodology provides a framework for more informed decision-making in agricultural experimental design and has potential applications in various studies focusing on optimal nitrogen application and site-specific fertilizer utilization.

By enabling more precise and efficient nitrogen management, our approach contributes to better crop yields, reduced environmental impact, and optimized resource utilization. This aligns with global goals for sustainability and food security, offering a valuable contribution to agricultural research and paving the way for future advancements in sustainable farming practices.

References

Alhorn, K., Dette, H., & Schorning, K. (2021). Optimal Designs for Model Averaging in non-nested Models. *Sankhya A*, *83*(2), 745-778. https://doi.org/10.1007/s13171-020-00238-9

Bachmaier, M. (2012). Sources of inaccuracy when estimating economically optimum N fertilizer rates. *Agricultural Sciences*, *Vol.03No.03*, 8, Article 19030. https://doi.org/10.4236/as.2012.33037

Bachmaier, M., & Gandorfer, M. (2009). A conceptual framework for judging the precision agriculture hypothesis with regard to site-specific nitrogen application. *Precision Agriculture*, *10*(2), 95-110. https://doi.org/10.1007/s11119-008-9069-x

Büchse, A., Gabriel, D., Grosse, M., Hartung, J., Kurz, S., Laidig, F., . . . Wittenburg, D. (2022). One, two, three: portable sample size in agricultural research. *The Journal of Agricultural Science*, *160*(6), 459-482. https://doi.org/10.1017/S0021859622000466

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846-4872. https://doi.org/https://doi.org/10.1016/j.ces.2007.11.034

Gilmour, S. G., & Trinca, L. A. (2012). Optimum Design of Experiments for Statistical Inference.

Journal of the Royal Statistical Society Series C: Applied Statistics, *61*(3), 345-401. https://doi.org/10.1111/j.1467-9876.2011.01000.x

López C, D. C., Barz, T., Körkel, S., & Wozny, G. (2015). Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. *Computers & Chemical Engineering*, *77*, 24-42. https://doi.org/https://doi.org/10.1016/j.compchemeng.2015.03.002

Meyer-Aurich, A., Griffin, T. W., Herbst, R., Giebel, A., & Muhammad, N. (2010). Spatial econometric analysis of a field-scale site-specific nitrogen fertilizer experiment on wheat (Triticum aestuvum L.) yield and quality. *Computers and Electronics in Agriculture*, *74*(1), 73-79. https://doi.org/https://doi.org/10.1016/j.compag.2010.06.007

Moles, C. G., Mendes, P., & Banga, J. R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome research*, *13*(11), 2467-2474.

Ng'ombe, J. N., & Brorsen, B. W. (2022). Bayesian optimal dynamic sampling procedures for on-farm field experimentation. *Precision Agriculture*, *23*(6), 2289-2311. https://doi.org/10.1007/s11119-022-09921-3

Pannell, D. J. (1997). Sensitivity analysis of normative economic models: theoretical framework and practical strategies. *Agricultural Economics*, *16*(2), 139-152. https://doi.org/https://doi.org/10.1016/S0169-5150(96)01217-0

Piepho, H.-P., Richter, C., Spilke, J., Hartung, K., Kunick, A., & Thöle, H. (2011). Statistical aspects of on-farm experimentation. *Crop and Pasture Science*, *62*(9), 721-735. https://doi.org/10.1071/CP11175

Poursina, D., & Brorsen, B. W. (2023). Fully Bayesian economically optimal design for a spatially varying coefficient linear stochastic plateau model over multiple years. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-023-02615-w

Poursina, D., Brorsen, B. W., & Lambert, D. M. (2023). Nearly Ds-optimal assigned location design for a linear model with spatially varying coefficients. *Spatial Statistics*, *53*, 100727. https://doi.org/https://doi.org/10.1016/j.spasta.2023.100727

Poursina, D., Brorsen, B. W., & Lambert, D. M. (2024). Optimal treatment placement for on-farm experiments: pseudo-Bayesian optimal designs with a linear response plateau model. *Precision Agriculture*. https://doi.org/10.1007/s11119-023-10105-w

Sharpley, A. N., Smith, S. J., & Naney, J. W. (1987). Environmental impact of agricultural nitrogen and phosphorus use. *Journal of Agricultural and Food Chemistry*, *35*(5), 812-817. https://doi.org/10.1021/jf00077a043

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C. M., & Gujer, W. (2009). Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, *43*(11), 2894-2906. https://doi.org/https://doi.org/10.1016/j.watres.2009.03.048

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.