

Bagging as a model averaging technique for estimating
optimum nitrogen rates in agricultural crops

by

Francisco Palmero

B.S., Universidad Nacional de Córdoba, 2019
M.S., Universidad de Buenos Aires, 2023
Ph.D., Kansas State University, 2024

A REPORT

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2025

Approved by:

Major Professor
Trevor J. Hefley

Copyright

© Francisco Palmero 2025.

Abstract

For several (non-legume) crops, the addition of nitrogen (N) via fertilization is key to maximizing grain yield and farmers' profit. The N rate balancing crop production and farmers' incomes is called the Economic Optimum N Rate (EONR). The process of finding the EONR is model-based, yielding different estimations depending on the data and the selected model. Therefore, in some cases, model averaging can be an appropriate tool to improve the accuracy (bias) and precision (variance) of the EONR estimation. In this study we propose for the first time the use of bagging, a simple machine learning algorithm, as a model averaging technique. The objectives of this study were to (i) evaluate the ability of bagging in estimating the EONR for different experimental designs (data quality) using simulated data, and (ii) apply this technique using data collected in a field experiment. In the simulation studies, we found that (beyond the model representing the true data generating process) bagging was able to reduce both bias and variance of the EONR estimation in comparison to individual models. This was more notorious as the number of observations and N rates increased (i.e. good data quality). For the real data example, bagging point estimation of the EONR was in between of the estimation of the individual models. Furthermore, bagging provided the wider confidence intervals of the EONR. Our framework for bagging grain yield response to N fertilization models provides researchers with a simple method to the estimation of the EONR (or other derived quantities), by considering a set of suitable candidate models.

Table of Contents

Table of Contents	iv
List of Figures	v
List of Tables	vi
Acknowledgements	vii
Dedication	viii
1) Introduction	1
2) Methods	3
2.1) Set of considered models and EONR estimation.....	3
2.2) Bagging.....	5
2.3) Data Simulation	8
2.4) Model performance evaluations	9
2.5) Case study	10
3) Results	11
3.1) Simulation results	11
3.2) Data example	19
4) Discussion.....	20
5) Conclusion.....	24
References	26
Appendix A.....	30
Supplementary Note 1	30
Appendix B	32
Supplementary Figures	32
Appendix C	34
R Code for Bagging	34

List of Figures

Figure 1. Flowchart depicting the bagging framework implemented in this study.	8
Figure 2. A) Relationship between bagging bias and bias averaged over individual models on the estimations of the EONR. The equation in the upper-left corner and the solid line represent the estimated expected value of the bagging bias. B) Relationship between bagging variance and variance averaged over individual models on the estimations of the EONR. The equation in the upper-left corner and the solid line indicate the estimated expected value of the bagging variance. In the reference for the different scenarios LP: the linear-plateau model was the data generating process, MIT: the Mitscherlich model was the data generating process, QD: the quadratic model was the data generating process, QP: the quadratic-plateau model was the data generating process, and “high”, “interm”, and “low” indicates high, intermediate, and low data quality respectively.	18
Figure 3. Relationship between maize grain yield and nitrogen rate. Circles indicates point estimation of the EONR and horizontal lines represents the 95% confidence interval of the EONR estimation.	20

List of Tables

Table 1. Results from simulation experiments when the quadratic model was the data generating process. $E[\theta]$ indicates the estimated EONR, $\text{Bias}[\theta]$ and $\text{Var}[\theta]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.	12
Table 2. Results from simulation experiments when the quadratic plateau model was the data generating process. $E[\theta]$ indicates the estimated EONR, $\text{Bias}[\theta]$ and $\text{Var}[\theta]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.	13
Table 3. Results from simulation experiments when the linear plateau model was the data generating process. $E[\theta]$ indicates the estimated EONR, $\text{Bias}[\theta]$ and $\text{Var}[\theta]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.	15
Table 4. Results from simulation experiments when the Mitscherlich model was the data generating process. $E[\theta]$ indicates the estimated EONR, $\text{Bias}[\theta]$ and $\text{Var}[\theta]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.	16

Acknowledgements

I want to thank Dr. Trevor Hefley for encouraging me in making this journey possible. I deeply appreciate Trevor's support, time, work, and the several pieces of conversations that we had about statistics and science in general. I also want to say thank you to Dr. Christopher Vahl and Dr. Perla Reyes for your contributions to this study.

I would like to express my gratitude to Dr. Ignacio Ciampitti, your support made this objective possible. I will be always grateful to you.

Special thanks to all my family. Thank you to my parents, Javier and Mónica for giving me all their support to being taking this journey.

Special thanks to the quasi-agronomist and statistician Caro, my life partner. Your support and patience are invaluable.

Dedication

To tulipa, chango, buena, vos decís que no, negra, lolo, and tintintin.

1) Introduction

Nitrogen (N) plays a major role in determining the performance of a crop. For several (non-legume) crops, the addition of N via fertilization is key to maximizing grain yield and farmers' profit (Bélangier et al., 2000; Jankowski et al., 2024; van Grinsven et al., 2022). Therefore, N fertilization is needed to match food production with food demand in a growing human population (Sinclair and Rufty, 2012). The practice of applying N to the soil is inefficient. A high proportion of the applied N is lost to the environment in aqueous or gaseous forms (Cassman et al., 2002). Thus, excessive N fertilizer applications cause negative environmental impacts (Galloway et al., 2003). However, N applications below the optimum decreases crop productivity affecting farmers' profit (Banger et al., 2020). Therefore, N rates should be estimated as accurate and precise as possible.

The relationship between crop grain yield and N rate is essential to improve N fertilization management considering the balance between economic return and environmental costs (Wang et al., 2014). Considering this relationship and the crop and N fertilizer prices, the Economic Optimum N Rate (EONR) can be estimated. This N rate is recommended to farmers with the aim of achieving maximum yield while considering economic and environmental efficiency (Sieling et al., 2023). Although the N recommendation system seeks to precisely and accurately estimate the EONR, this is a challenging task due to the presence of different sources of uncertainty.

There are two big groups causing uncertainty on the estimation of the EONR. One of them encompasses "natural" causes such as changes in weather (e.g., temperature, precipitation), soil N mineralization and availability, and crop growth and N demand (Briat et al., 2020; Lobell, 2007). The other source is related to the analysis of the data. Estimating the EONR is a model

based-approach. Therefore, the inference about the EONR lies on the validity of the statistical model selected to describe the relationship between grain yield and N fertilizer rate. Several mathematical models are considered appropriate by scientists to model crop grain yield response to N fertilizer application (Kyveryga et al., 2007; Lyons et al., 2019). Defining which functional form or model to use is both user and data dependent, generating a new source of uncertainty related to model selection.

In maize (*Zea mays*, L), different studies have shown variation on the estimation of the EONR depending on the used model (Morris et al., 2018). Each of the models can be considered as a hypothesis explaining the data generating process. Then, we are considering multiple hypothesis simultaneously lying in the framework of multimodel inference. One of the most widely used multimodel inference technique is the implementation of information criteria to select the “best” model according to the data. The most known and used information criteria are Aikake’s (AIC; Akaike, 1998) and Schwartz or Bayesian (BIC; Schwarz, 1978). However, the practice of choosing the “best” model based on information criteria is not always clear (Miguez and Poffenbarger, 2022) and might be the case that the “best” model is not a good representation of the data (Mac Nally et al., 2018).

The selection of a single “best” model might imply considerable under- or over-estimation of the EONR (Jaynes, 2011; Kyveryga et al., 2007). Therefore, making an average of different models can be a suitable multimodel inference technique for obtaining accurate and precise predictions of the EONR (Burnham and Anderson, 2002). Miguez and Poffenbarger (2022) and Matavel et al. (2024) evaluated the use of model averaging to estimate the EONR based on AIC model weights. In this study, we introduce a simple machine learning algorithm

called bootstrapping aggregation or “bagging” (Breiman, 1996) as a technique to perform model averaging to estimate the EONR.

The objectives of this study were to (i) evaluate the ability of bagging in estimating the EONR for different experimental designs (data quality) using simulated data, and (ii) apply this technique using data collected in a field experiment. The applied bagging algorithm is explained in detail in the materials and methods section. In this study we focus on the maize grain yield response to N. However, this methodology is equally useful in other crops and analyses such as estimating optimum seeding rate and model predictions.

2) Methods

2.1) Set of considered models and EONR estimation

Several mathematical models can be implemented to model the expected value of grain yield as function of N rates (Lyons et al., 2019). For this particular study we will consider a set of 4 statistical models, $M = \{M_1, \dots, M_4\}$. The selected models are defined based on expert scientific knowledge and the assumptions behind them are not evidently inappropriate. For all the models in this study, it was considered that $\mathbf{y} \sim normal(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$, where $\mathbf{y}$ is the vector of observations of grain yield and σ^2 is the variance of grain yield assuming to be the same for all y_i 's in $\mathbf{y}$. The main difference among models was the mathematical equation used to model the expected value of each normal probability function, i.e. each μ_i in the vector $\boldsymbol{\mu}$. The mathematical equations for the different statistical models were:

Quadratic ($M = 1$)

$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \quad (1)$$

Quadratic plateau ($M = 2$)

$$\mu_i = \begin{cases} \beta_0 + \beta_1 x_i + \beta_2 x_i^2 & \text{if } x_i < AONR \\ \beta_0 + \beta_1 AONR + \beta_2 AONR^2 & \text{o.w.} \end{cases} \quad (2)$$

where AONR is the agronomic optimum N rate obtained as $AONR = \frac{\beta_1}{2\beta_2}$. This value indicates the N rate at which corn achieves the maximum attainable yield.

Linear plateau ($M = 3$)

$$\mu_i = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_i < xp \\ \beta_0 + \beta_1 xp & \text{o.w.} \end{cases} \quad (3)$$

where xp is the N rate at which the grain yield becomes independent of further N inputs staying at the maximum attainable yield.

Mitscherlich ($M = 4$)

$$\mu_i = \beta_0 + (\beta_1 - \beta_0) e^{-\beta_2 x_i}. \quad (4)$$

In all of these equations, x_i represents the N rate associated with the i th observation of grain yield.

The parameters obtained in equations (1) to (4) can be implemented to calculate the EONR. Since the EONR is the N rate at which the net return to farmers is maximized, the cost of the N fertilizer and the price of the crop are considered in the calculation of the EONR. All the functions implemented to model the expected values above, except the linear plateau, have a first derivative where x , i.e. the N rate, is involved. Then, for those models the first derivative of μ can be used for obtaining the EONR. In those cases, the first derivative is equated to the ratio between crop price and N price, and the equations are solved for the N rate (x). For the quadratic and quadratic plateau model this is

$$EONR = \frac{\beta_1 - PR}{2\beta_2}, \quad (5)$$

where $PR = \frac{\text{Corn price } (\$/kg)}{\text{N price } (\$/kg)}$, which indicates how many kilos of corn are needed to pay for one

kilo of N in the fertilizer. For the Mitscherlich model,

$$EONR = -\frac{\ln \left[\frac{-p_f}{p_c(\beta_1 - \beta_0)\beta_2} \right]}{\beta_2}, \quad (6)$$

where p_f and p_c are the N fertilizer cost and corn price, respectively. For the linear plateau function, if $p_f > p_c\beta_1$, then $EONR = 0$, otherwise $EONR > 0$. In our study, in all cases $p_f < p_c\beta_1$ meaning that the parameter xp in (3) was considered to be the EONR for the linear plateau model. When computing the EONR for quadratic, quadratic plateau, and Mitscherlich models, corn price = 0.789 \$/kg and N price = 0.154 \$/kg, which are the average prices over 2000-2022 years in United States (USDA ERS, 2024; USDA ERS Price Forecast, 2024).

2.2) Bagging

Bagging is an ensemble machine learning algorithm designed to improve the precision and stability of statistical classification and regression problems (Breiman, 1996). What does “ensemble” mean? Ensemble means that the implemented technique uses two or more algorithms (e.g. regression models, regression trees, neural networks, etc.) to learn about relationships among different variables producing more precise predictions than a single model (Zhang and Ma, 2012). For example, one of the most used and known ensemble techniques in agriculture studies is random forest, which is an ensemble of regression or decision trees models (Breiman, 2001). As an ensemble method, bagging consists of training the models over several datasets and then combine them to obtain a compound estimation or prediction with lower variance (Breiman, 1996; Zhang and Ma, 2012). To achieve this objective, bagging is split into two parts: bootstrapping and aggregation.

Bootstrapping is a sampling method to take random samples from the original data set with replacement (Efron and Tibshirani, 1986). Sampling with replacement allows us to ensure that once we take a sample the subsequent samples are independent from the previous. The first part of the bagging technique uses bootstrapping to build K new training sets by taking random samples (with replacement) from the original data. Each of the K training sets is of the same size as the original data (n). Before jumping to the second part of bagging (i.e. aggregation), the learning algorithms (e.g. regression models, k-nearest neighbor regression, etc.) are run on the new training sets to obtain the parameter estimation or prediction. Therefore, at the end of this stage, we end up with K parameter estimations or predictions each one coming from different learning algorithms. The second part, aggregation, simply means to combine the previously obtained estimations or predictions to derive a single one. It is expected that the estimation or prediction derived from aggregation has lower variance than the one coming from the individual learning algorithms.

Now, we will explain the application of bagging to the particular case of getting more precise estimations of the EONR. For this case, the different learning algorithms are the regression models explained in (1) to (4). Consider the collected data to be stored in a matrix

$$\mathbf{D} = \begin{bmatrix} y_1 & x_1 \\ \vdots & \vdots \\ y_n & x_n \end{bmatrix}, \text{ where } y_1, \dots, y_n \text{ are the observed grain yields and the } x_1, \dots, x_n \text{ is the associated}$$

N rate. This matrix has n rows and 2 columns, then it is a $n \times 2$ matrix, where n is the total number of observations in the original data set. The applied bagging procedure consisted of six steps that were represented in Figure 1. These steps were:

(i) obtain $K=5000$ new matrices of size $n \times 2$ using bootstrapping. These matrices are represented as $\mathbf{D}^{(k)}$, where the superindex k indicates the k th new training set.

- (ii) Randomly select one of the candidate models with probability $1/L$, where L is the total number of candidate models.
- (iii) Fit the selected model to the data in $\mathbf{D}^{(k)}$, i.e. to the k th new training set.
- (iv) Compute the EONR according to the model fitted in (iii).
- (v) Save the estimated EONR.
- (vi) Repeat steps (i) to (v) until a given candidate model was fitted to each of the k th new training sets.

Once this algorithm was applied, we ended up with $K=5000$ estimations of the EONR coming from different learners or candidate models. These K estimations formed an empirical probability distribution of the EONR, which was used to obtain point and uncertainty estimations of this quantity of interest. The point estimation can be performed by taking the median or the mean of the empirical distribution. Here is where the aggregation step of the bagging procedure occurred. The uncertainty estimation can be performed by computing the standard deviation or given quantiles of the empirical distribution. In our case, we performed point estimation by taking the median and the uncertainty estimation by considering the quantiles 0.025 and 0.975 of the empirical distribution to obtain the 95% confidence interval. The step (iv) can be adapted to obtain the estimation of other derived quantities such as the AONR or the expected grain yield at a given N rate.

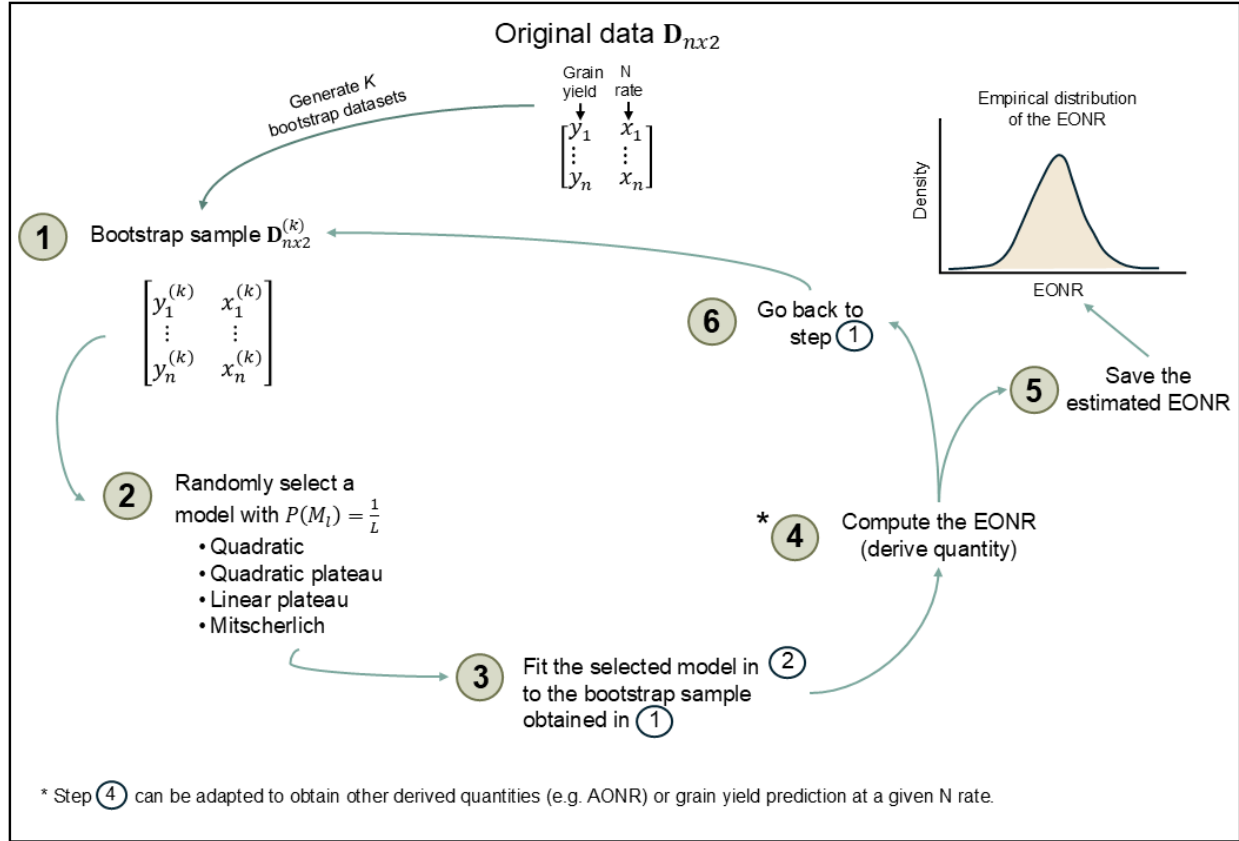


Figure 1. Flowchart depicting the bagging framework implemented in this study.

2.3) Data Simulation

Simulating data allowed us to know the true value of the model parameters and the EONR. Therefore, by simulating data we can evaluate the performance of bagging and individual models in terms of their inference properties (e.g. bias and variance). For this exercise, we considered different scenarios of data quality and data generating processes. According to Supplementary Figure 1 in Miguez and Poffenbarger (2022), we defined three types of data quality determined by the standard error on the estimation of the agronomic optimum N rate (a similar quantity than the EONR). The data types were: (i) low quality: the data was comprised by 5 N rates and 3 repetitions, (ii) intermediate quality: 7 N rates with 4 repetitions were considered, and (iii) high quality: the data consisted of 9 N rates with 6 repetitions. The total number of

observations were 15, 28, and 54 for the low, intermediate, and high data quality cases, respectively. Furthermore, for all cases the minimum N rate was 0 kg ha⁻¹ and the maximum it was considered 315 kg ha⁻¹. Thus, the increments between N rates were 78.75 kg ha⁻¹ (low quality), 52.50 kg ha⁻¹ (intermediate quality), and 39.37 kg ha⁻¹ (high quality). For the three data quality, the data was simulated considering a completely randomized design.

For each data quality case, it was assumed that $y_i \sim normal(\mu_i, \sigma)$, where y_i was corn grain yield and μ_i and σ were the expected value and the standard deviation of grain yield, respectively. The parameter σ was set to 1000 kg ha⁻¹. The equations in (1) to (4) were considered the true data generating processes determining μ_i . The models in (1) to (4) to the grain yield response to N were fitted to the experiments used in Correndo et al. (2021). The obtained estimated parameters were considered as “true” parameters and used for simulation. For all the simulated data sets in each data quality condition, the true EONR was 199.3, 197.8, 145.3, and 222.2 kg ha⁻¹ for the quadratic, quadratic plateau, linear plateau, and Mitscherlich models, respectively. Considering the three data quality types and the four true data generating processes, a total of 12 scenarios were considered. In each scenario a total of 500 data sets were simulated. A seed was set for reproducibility. Therefore, bagging and individual models were fitted to the same 500 simulated data sets in each scenario. Examples of simulated data for all the scenarios are shown in Supp. Figure 1. The simulations were carried out in R software (R Core Team, 2020).

2.4) Model performance evaluations

Bagging and individual models were used for estimating the EONR in the 500 data sets for each of the 12 scenarios. The models that did not converge were removed from the final set

of EONR estimations. A commonly used criterion to evaluate estimators in classical statistics is the mean squared error or MSE. Let W be the total number of simulated data sets (i.e. $W=500$), θ the true EONR, and $\hat{\theta}_{i(mj)}$ be the estimated EONR for the i th data set using the m th model in the j th scenario. In this case, the ensemble model obtained by bagging is considered in the set, then $m = 1, \dots, 5$ and $j = 1, \dots, 12$. Therefore,

$$MSE_{mj} = Var \hat{\theta}_{mj} + (Bias \hat{\theta}_{mj})^2, \quad (7)$$

where

$$Var \hat{\theta}_{mj} = \frac{1}{W-1} \sum_{i=1}^W \left[\hat{\theta}_{i(mj)} - \left(\frac{1}{W} \sum_{i=1}^W \hat{\theta}_{i(mj)} \right) \right]^2, \quad (8)$$

and

$$Bias \hat{\theta}_{mj} = \left(\frac{1}{W} \sum_{i=1}^W \hat{\theta}_{i(mj)} \right) - \theta. \quad (9)$$

Therefore, the MSE has two components, the variance measuring the variability of the estimator (i.e. precision) and the bias measuring its accuracy. An estimator that has good MSE properties has small, combined variance and bias.

2.5) Case study

The open dataset of maize grain yield response to N presented in Ransom et al. (2021) was utilized to apply the evaluated method to real experimental data. The trial number 36 carried out in Urbana, Illinois, in season 2015 was randomly selected. The models presented in (1) to (4) were fitted to the data individually and via bagging. The confidence intervals of the estimated EONR were computed. For the individual models, it was assumed that $\hat{\theta}$ was normally distributed and the confidence intervals were estimated as

$$CI(\theta_m) = \hat{\theta}_m \pm 1.96\sqrt{\widehat{Var}(\hat{\theta}_m)},$$

where $\hat{\theta}_m$ is the EONR estimation of the m th model. For the quadratic, quadratic plateau, and Mitscherlich models, the variance of the EONR estimation, i.e. $\widehat{Var}(\hat{\theta}_m)$, was approximated by using the delta method (Ver Hoef, 2012; Supp. Note 1). For the linear plateau model, the EONR was an estimated model parameter, then $\widehat{Var}(\hat{\theta}_m)$ was obtained from the asymptotic covariance matrix of the model parameters using the Jacobian matrix of the model (Bates and Watts, 2007; Supp. Note 1). For bagging, the confidence intervals were computed as it was described in section 2.2. All the analyses were performed in R software (R Core Team, 2020). The *nls()* function was used to fit the non-linear models to the data and the package *msm* was used to apply the delta method (Jackson, 2011).

3) Results

3.1) Simulation results

We conducted simulation studies in twelve scenarios determined by the combination of three different data quality and four true data generating processes. For each scenario, we evaluated the bias and variance of the individual models and of bagging as a model averaging technique across 500 simulated datasets. When the true data generating process was the quadratic model, the bias and variance decompositions are presented in Table 1. Under this context the quadratic model showed the lowest MSE in all the three data quality conditions. Bagging achieved the smallest MSE among the rest of the models, even in the low data quality scenario. The smaller MSE of bagging was achieved by a reduction in bias and variance (in comparison with the quadratic plateau model), bias (with respect to the linear plateau model), or in variance (in contrast to the Mitscherlich model) (Table 1).

Table 1. Results from simulation experiments when the quadratic model was the data generating process. $E[\hat{\theta}]$ indicates the estimated EONR, $\text{Bias}[\hat{\theta}]$ and $\text{Var}[\hat{\theta}]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.

Data Quality	Model	$E[\hat{\theta}]$	$\text{Bias}[\hat{\theta}]$	$\text{Var}[\hat{\theta}]$	MSE
Low (n=15)	Bagging	183.2	-16.1	585.6	843.8
	Linear Plateau	127.8	-71.5	463.8	5573.9
	Mitscherlich	191.2	-8.1	1773.7	1839.7
	Quadratic	201.9	2.6	246.7	253.6
	Quadratic Plateau	174.8	-24.5	1036.1	1636.8
Intermediate (n=28)	Bagging	184.2	-15.0	364.2	589.8
	Linear Plateau	131.9	-67.4	351.0	4896.7
	Mitscherlich	199.6	0.3	809.7	809.8
	Quadratic	200.3	1.0	131.6	132.6
	Quadratic Plateau	177.8	-21.5	543.4	1005.6
High (n=52)	Bagging	187.5	-11.8	206.4	345.9
	Linear Plateau	136.1	-63.2	167.8	4155.8
	Mitscherlich	203.1	3.8	443.7	458.2
	Quadratic	200.5	1.2	67.5	69.0
	Quadratic Plateau	179.7	-19.5	301.9	684.0

Table 2. Results from simulation experiments when the quadratic plateau model was the data generating process. $E[\hat{\theta}]$ indicates the estimated EONR, $\text{Bias}[\hat{\theta}]$ and $\text{Var}[\hat{\theta}]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.

Data Quality	Model	$E[\hat{\theta}]$	$\text{Bias}[\hat{\theta}]$	$\text{Var}[\hat{\theta}]$	MSE
Low (n=15)	Bagging	209.8	12.0	1002.3	1146.3
	Linear Plateau	135.4	-62.4	570.4	4464.1
	Mitscherlich	240.9	43.1	4014.4	5875.9
	Quadratic	229.0	31.2	883.2	1855.4
	Quadratic Plateau	199.6	1.8	1958.9	1962.2
Intermediate (n=28)	Bagging	209.9	12.1	518.5	665.0
	Linear Plateau	140.4	-57.4	400.0	3690.5
	Mitscherlich	243.3	45.5	1583.9	3658.5
	Quadratic	225.4	27.6	291.2	1055.6
	Quadratic Plateau	199.9	2.1	837.8	842.2
High (n=52)	Bagging	209.5	11.7	261.4	398.7
	Linear Plateau	143.5	-54.3	186.8	3137.0
	Mitscherlich	239.4	41.6	701.9	2432.4
	Quadratic	221.5	23.7	147.7	710.3
	Quadratic Plateau	197.8	0.1	397.4	397.4

In Table 2, we report the bias and variance of bagging and the individual models when the true data generating process was the quadratic plateau model. In this case, the quadratic plateau model was the one with the smallest bias, but bagging was the technique with the smallest MSE, across all data quality scenarios. When the linear plateau model was the true data generating process, the performance of bagging was not superior to that of the linear and quadratic plateau models (Table 3). In this case, all models beyond the linear plateau had a

considerable higher bias than in the other cases, leading bagging to a poor overall performance. In the case of the Mitscherlich model being the true data generating process, the quadratic and Mitscherlich had the lowest bias (Table 4). However, bagging achieved a smaller variance than the Mitscherlich model having the second smallest MSE. As a summary, across the twelve scenarios studied, bagging was the technique with the best balance among bias and variance, i.e. lowest MSE, after the true data generating model.

Table 3. Results from simulation experiments when the linear plateau model was the data generating process. $E[\hat{\theta}]$ indicates the estimated EONR, $\text{Bias}[\hat{\theta}]$ and $\text{Var}[\hat{\theta}]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.

Data Quality	Model	$E[\hat{\theta}]$	$\text{Bias}[\hat{\theta}]$	$\text{Var}[\hat{\theta}]$	MSE
Low (n=15)	Bagging	207.8	62.5	829.5	4742.3
	Linear Plateau	142.9	-2.3	592.2	597.5
	Mitscherlich	241.7	95.5	3413.7	12716.6
	Quadratic	224.0	78.7	704.8	6409.7
	Quadratic Plateau	199.3	54.0	1355.8	4272.7
Intermediate (n=28)	Bagging	209.2	63.9	435.9	4519.7
	Linear Plateau	144.7	-0.5	292.1	292.4
	Mitscherlich	246.0	100.7	1436.5	11585.0
	Quadratic	221.6	76.4	279.1	6111.7
	Quadratic Plateau	199.4	54.1	648.4	3575.9
High (n=52)	Bagging	207.7	62.4	204.3	4101.9
	Linear Plateau	145.0	-0.2	138.9	138.9
	Mitscherlich	241.2	96.0	641.3	9851.1
	Quadratic	217.8	72.5	123.0	5384.1
	Quadratic Plateau	195.2	50.0	292.0	2790.0

Table 4. Results from simulation experiments when the Mitscherlich model was the data generating process. $E[\hat{\theta}]$ indicates the estimated EONR, $\text{Bias}[\hat{\theta}]$ and $\text{Var}[\hat{\theta}]$ represents the average bias and the variance on the EONR estimation across the 500 datasets. MSE is the Mean Squared Error.

Data Quality	Model	$E[\hat{\theta}]$	$\text{Bias}[\hat{\theta}]$	$\text{Var}[\hat{\theta}]$	MSE
Low (n=15)	Bagging	209.1	-13.1	1036.2	1207.6
	Linear Plateau	125.6	-96.6	460.9	9790.5
	Mitscherlich	232.6	10.4	3871.7	3980.1
	Quadratic	232.7	10.5	806.0	916.6
	Quadratic Plateau	193.0	-29.2	2216.2	3071.1
Intermediate (n=28)	Bagging	202.5	-19.7	731.1	1118.1
	Linear Plateau	129.3	-93.0	350.2	8992.0
	Mitscherlich	226.7	4.5	1526.0	1546.1
	Quadratic	226.3	4.1	355.2	371.6
	Quadratic Plateau	188.9	-33.3	1282.9	2392.3
High (n=52)	Bagging	203.9	-18.3	430.0	765.6
	Linear Plateau	133.9	-88.3	198.6	7997.0
	Mitscherlich	225.4	3.2	777.5	787.9
	Quadratic	224.8	2.5	164.5	170.9
	Quadratic Plateau	188.7	-33.5	696.3	1821.8

Considering the twelve scenarios, independently of the data generating process, the bias and variance, of individual models and bagging decreased as the data quality (sample size and number of N rates) increased. In Figure 2A, we depict the relationship between bagging bias and the average bias over individual models. The slope of the simple regression model was 0.962, indicating that a unit increment, on average, in the bias of the individual models bagging bias increased in one unit. However, in Figure 2B it is shown that a unit increment, on average, in the variance of the individual models, the variance of bagging increased in less than one unit (slope = 0.49). These results highlight two main points, (i) the performance of bagging was affected by the performance of the individual models, and (ii) the variance of bagging was less affected by the individual models than the bias. Therefore, bagging worked mainly as a variance stabilization technique.

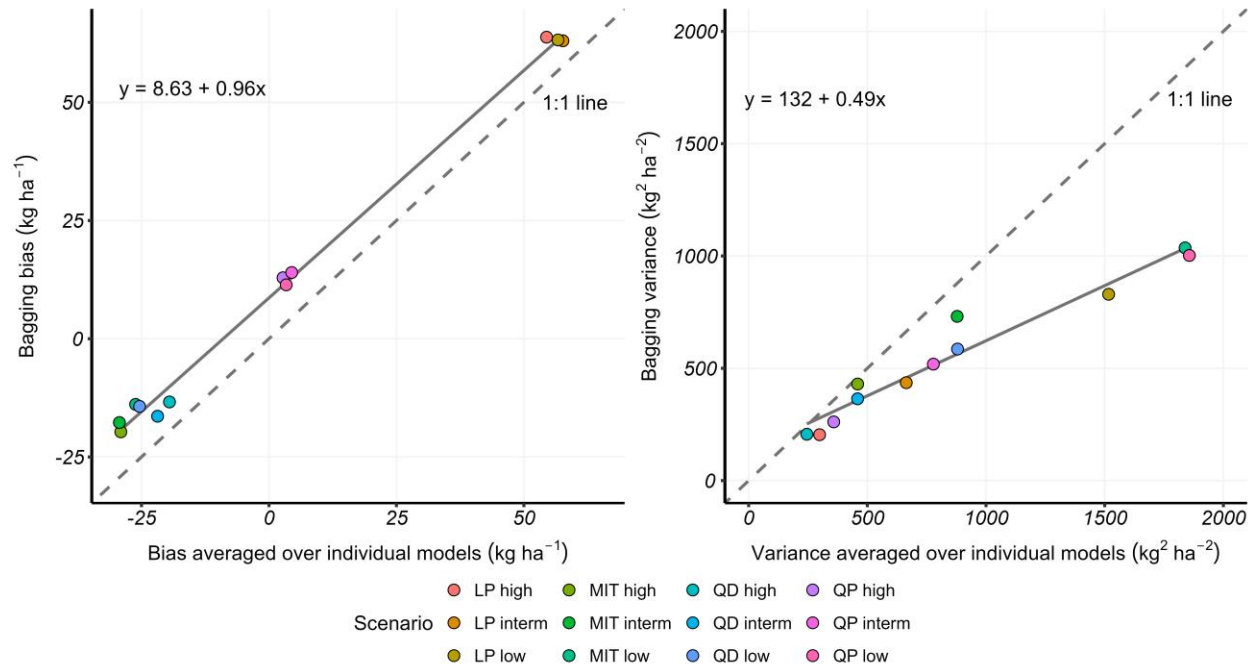


Figure 2. A) Relationship between bagging bias and bias averaged over individual models on the estimations of the EONR. The equation in the upper-left corner and the solid line represents the estimated expected value of the bagging bias using least square estimation. B) Relationship between bagging variance and variance averaged over individual models on the estimations of the EONR. The equation in the upper-left corner and the solid line indicates the estimated expected value of the bagging variance using least square estimation. In the reference for the different scenarios LP: the linear-plateau model was the data generating process, MIT: the Mitscherlich model was the data generating process, QD: the quadratic model was the data generating process, QP: the quadratic-plateau model was the data generating process, and “high”, “interm”, and “low” indicates high, intermediate, and low data quality respectively.

Once bagging is applied, there are still uncertainty on the estimation of the EONR.

Therefore, the confidence intervals generated by bagging were analyzed. In Supp. Figure 2, we show the confidence intervals obtained by using bagging for the 500 simulated data sets in the twelve scenarios. The results showed that for low data quality (small sample sizes and number of N rates), the uncertainty obtained by bagging led to ill-constrained upper bounds (Supp. Figure 2). Therefore, with low data quality, bagging might lead to confidence intervals that lack agronomic meaning. This problem tended to disappear as the sample size and the number of N rates increased (Supp. Figure 2).

3.2) Data example

In Figure 3, we present the results of applying bagging and fitting the individual models to a single experimental maize grain yield response to N dataset. The sample size and the number of N rates were 64 and 7, respectively. The individual models provided different EONR estimations. The linear plateau model produced the lowest estimation of 146 kg ha^{-1} . On the other hand, the Mitscherlich model estimated the EONR to be 239 kg ha^{-1} . The bagging point estimation of the EONR was 212 kg ha^{-1} being in between the individual model estimations (Figure 3). The range of the confidence intervals obtained by the individual models was from 24 kg ha^{-1} to 69 kg ha^{-1} , for the linear plateau and Mitscherlich models, respectively. The range of the confidence interval yielded by bagging was 131 kg ha^{-1} , achieving wider confidence interval estimation than the individual models.

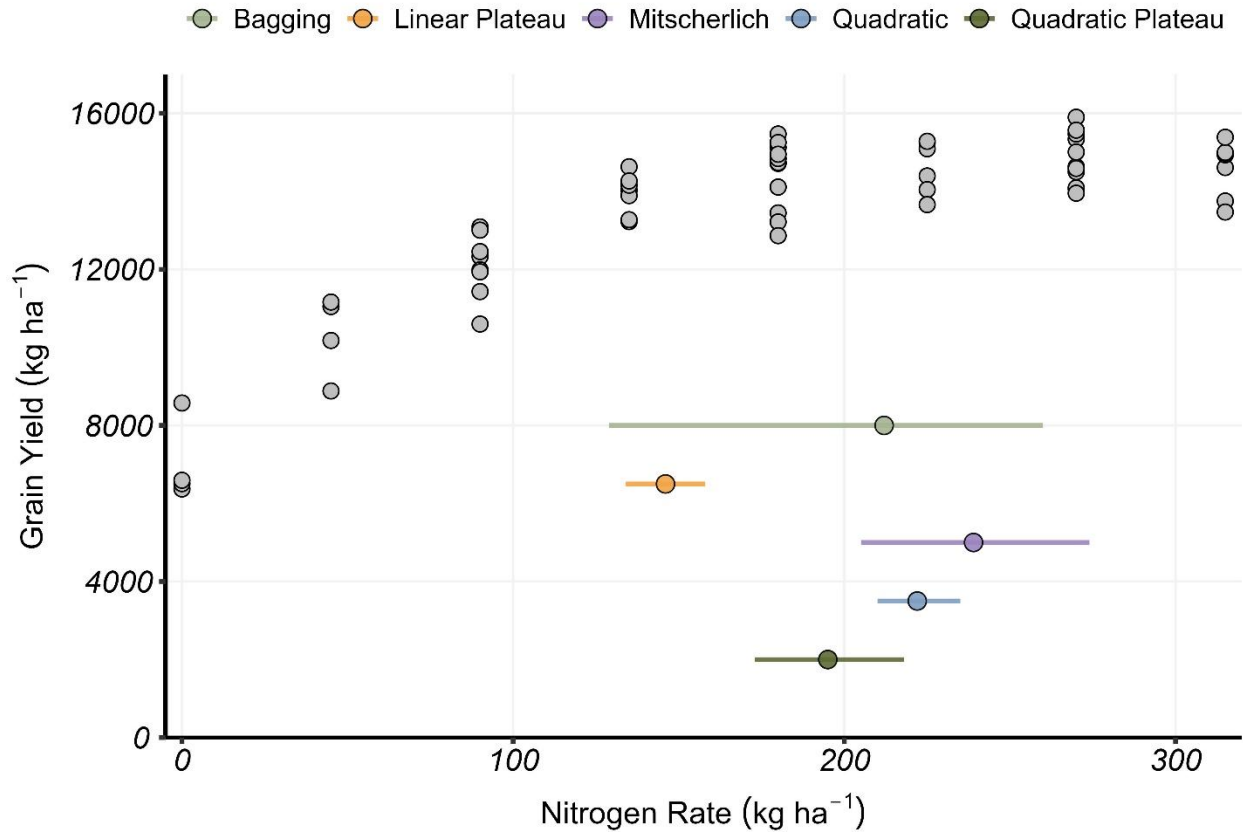


Figure 3. Relationship between maize grain yield and nitrogen rate. Circles indicate point estimation of the EONR and horizontal lines represents the 95% confidence interval of the EONR estimation.

4) Discussion

In this study, we introduced bagging, a simple machine learning algorithm, as a multimodel inference technique to estimate the EONR in agricultural crops. Previous studies have shown that model averaging (using weighted models) was an appropriate technique to estimate the EONR by effectively addressing the uncertainty associated with model specification (Matavel et al., 2024; Miguez and Poffenbarger, 2022). Our framework for bagging grain yield response to N fertilization models provides researchers with a simple method to the estimation of derived quantities (such as the EONR), by considering a set of suitable candidate models.

Furthermore, the estimations of the EONR are used to build an empirical distribution facilitating

the computation of standard errors and confidence intervals of the EONR estimation. This allows researchers to avoid the use of relatively complicated formulas in comparison with obtaining the uncertainty estimation implementing model averaging via weighted models using AIC or BIC (Burnham and Anderson, 2002). The proposed bagging framework provides a new alternative to perform multimodel inference to estimate not only the EONR, but other derived quantities of interest where there are multiple candidate models.

Overall, beyond the model generating the data, bagging showed the best balance between bias and variance, leading to a small MSE even under small sample sizes. This was achieved by combining a set of trained models over different datasets (using bootstrapping) to obtain a compound estimation of the EONR with smaller variance than the individual models (Zhang and Ma, 2012). The performance of bagging was affected by the performance of the individual models included in the set (Kuncheva and Whitaker, 2003; Wood et al., 2023) as it was also shown for model averaging using AIC (Matavel et al., 2024). We demonstrated that bagging was more effective at reducing the variance than the bias of the EONR estimation (Bauer and Kohavi, 1999). Different techniques such as adding crossover operations in the algorithm (Ngo et al., 2022) or introducing a regularization term in the loss function (Ugirimurera et al., 2024) might be implemented to improve bagging performance in terms bias. However, it should be considered that the datasets implemented in grain yield response to N studies are usually small (Miguez and Poffenbarger, 2022) in comparison with the target datasets for machine learning techniques.

Bagging showed high MSE for estimating the EONR when the linear plateau model was the data generating process. This might be related to the fact that ensemble techniques work because different models will usually make different errors on new data points (Goodfellow et

al., 2016). The quadratic, quadratic plateau, and Mitscherlich models consistently overestimated the EONR by, on average, 75 kg ha^{-1} when the linear plateau model generated the data. This means that in this case, all the learners made similar errors leading bagging to have poor performance (Bühlmann and Yu, 2002). Therefore, including models showing more contrasting estimations of the EONR might help to improve the performance of bagging. The probability of including models having more accurate EONR estimations increases as the number of models included in the set increases. Therefore, having more candidate models in the set (that are not wrong in terms of assumptions) can improve the performance of bagging by increasing the diversity of the ensemble algorithm (Wood et al., 2023). However, adding more models in the set will not improve bagging performance indefinitely due to the trade-off between diversity and accuracy (Johnson and Giraud-Carrier, 2019).

We applied the proposed bagging framework to estimate the EONR in a single experimental dataset of maize grain yield response to N. We showed that different models yielded different EONR estimations (Morris et al., 2018). In practice, there are no known data-generating models. An option to address this problem might be to select the most appropriate model based on the data at hand using some information criteria. However, the selected model might not be the “best” model (Matavel et al., 2024; Miguez and Poffenbarger, 2022). Therefore, estimating the EONR using bagging is a reasonable choice when the expert knowledge does not allow the researcher to determine which model is the most appropriate in terms of assumptions (which include functional forms). In the data example, the wider confidence interval obtained with bagging compared to those of individual models might be due to two reasons: (i) bagging EONR estimation considered not only the uncertainty on the estimation but also the uncertainty in model selection, and (ii) the delta method usually tends to underestimate the standard errors in

comparison to bootstrapping (Efron, 1990). Although narrow confidence intervals indicate higher precision on the estimation, the accuracy is not guaranteed. Thus, when the data do not show a clear functional form, applying bagging might be a more cautious option in comparison to selecting one of the individual models.

In this study, we showed that bagging had a good performance as model averaging technique for estimating the EONR. In general, we accepted a relatively small amount of bias on the estimation of the EONR in return for a larger reduction in variance, leading to small MSE. Therefore, bagging can be used to reduce overfitting making the EONR estimation more stable across different datasets (Soloff et al., 2024). This is the main objective of any regularization technique such as Lasso and ridge regressions. Although ensemble methods do not introduce a penalty or constrain term in a loss function as the previously mentioned techniques do, these methods lie within the general framework of regularization by combining multiple models (Friedrich et al., 2023).

The point and confidence interval estimation of bagging improved as the sample size and the number of N rates increased. Under small number of observations and N rates, the EONR uncertainty estimation had poorly constrained upper bounds yielding values lack of agronomic sense. This might be related to the improved consistency and efficiency of the bootstrap technique under large sample sizes (Efron and Tibshirani, 1986). To overcome this issue, Bayesian model averaging might be applied. This technique would allow us to regularize (constrain) the EONR prediction by adding a regulator in form of informative priors for model parameters. Bayesian model averaging also combines multiple models (similar to bagging) which would help to obtain a reduction in variance making the EONR estimation more stable across different datasets (Friedrich et al., 2023). Furthermore, there are several candidate models

to represent the grain yield response of crops to N fertilization. We showed that bagging was affected by the performance of the individual models, mainly in terms of bias. Therefore, future studies should consider more diverse models and consider diversity as a hidden dimension in the variance-bias decomposition of the ensemble (Wood et al., 2023).

We showed that bagging is an appropriate technique to obtain a reasonable EONR estimation when there are important mismatches among different individual models. However, although model averaging is reasonable and logical, it is not always desirable (Ver Hoef and Boveng, 2015). Models allow us, via simplification, to interpret and understand how a natural process works. This is the base of deductive-inductive reasoning in science (Howson and Urbach, 2005). Therefore, considering a single model by its diagnostic, interpretation, and correction is key to moving science forward. If the parameters of a single model have mechanistic basis and it is suitable for the objective of the study, then we can base our inference on that single model. However, if researchers are not satisfied by making inference conditional on a single model, then model averaging can relax this condition. Furthermore, if the final objective is just predicting some outcome, or simply we need more than one model to obtain insights from a particular problem, then model averaging can also be appropriate. What we say here is that model averaging is an statistical tool, that like all others, must be appropriately used (Banner and Higgs, 2017).

5) Conclusion

In this study we proposed bagging as a model averaging technique to estimate the EONR in agricultural crops. Overall, bagging showed higher accuracy and smaller variance than the individual models. Our proposed framework is simple and requires minimum knowledge of

machine learning algorithms and elementary statistics. Little adjustment is needed to adapt this framework to obtain other derived quantities such as the AONR, optimum seed rates, or yield predictions using a set of candidate models. In general, there is more than one suitable statistical approach to tackle a given question or problem. We believe that our simple bagging framework provides a simple new approach to the statistical toolbox of agricultural researchers.

References

- Akaike, H., 1998. Information Theory and an Extension of the Maximum Likelihood Principle, in: Parzen, E., Tanabe, K., Kitagawa, G. (Eds.), *Selected Papers of Hirotugu Akaike*. Springer, New York, NY, pp. 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15
- Banger, K., Nasielski, J., Janovicek, K., Sulik, J., Deen, B., 2020. Potential Farm-Level Economic Impact of Incorporating Environmental Costs Into Nitrogen Decision Making: A Case Study in Canadian Corn Production. *Front. Sustain. Food Syst.* 4. <https://doi.org/10.3389/fsufs.2020.00096>
- Banner, K.M., Higgs, M.D., 2017. Considerations for assessing model averaging of regression coefficients. *Ecological Applications* 27, 78–93. <https://doi.org/10.1002/eap.1419>
- Bates, D.M., Watts, D.G., 2007. *Nonlinear Regression Analysis and Its Applications*, 1st edition. ed. Wiley-Interscience, New York, NY.
- Bauer, E., Kohavi, R., 1999. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning* 36, 105–139. <https://doi.org/10.1023/A:1007515423169>
- Bélanger, G., Walsh, J.R., Richards, J.E., Milburn, P.H., Ziadi, N., 2000. Comparison of Three Statistical Models Describing Potato Yield Response to Nitrogen Fertilizer. *Agronomy Journal* 92, 902–908. <https://doi.org/10.2134/agronj2000.925902x>
- Breiman, L., 2001. Using Iterated Bagging to Debias Regressions. *Machine Learning* 45, 261–277. <https://doi.org/10.1023/A:1017934522171>
- Breiman, L., 1996. Bagging predictors. *Mach Learn* 24, 123–140. <https://doi.org/10.1007/BF00058655>
- Briat, J.-F., Gojon, A., Plassard, C., Rouached, H., Lemaire, G., 2020. Reappraisal of the central role of soil nutrient availability in nutrient management in light of recent advances in plant nutrition at crop and molecular levels. *European Journal of Agronomy* 116, 126069. <https://doi.org/10.1016/j.eja.2020.126069>
- Bühlmann, P., Yu, B., 2002. Analyzing bagging. *The Annals of Statistics* 30, 927–961. <https://doi.org/10.1214/aos/1031689014>
- Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information Theoretic Approach, in: *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, NY, pp. 437–454. https://doi.org/10.1007/978-0-387-22456-5_8
- Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, Nitrogen-use Efficiency, and Nitrogen Management. *ambi* 31, 132–140. <https://doi.org/10.1579/0044-7447-31.2.132>
- Correndo, A.A., Tremblay, N., Coulter, J.A., Ruiz-Diaz, D., Franzen, D., Nafziger, E., Prasad, V., Rosso, L.H.M., Steinke, K., Du, J., Messina, C.D., Ciampitti, I.A., 2021. Unraveling uncertainty drivers of the maize yield response to nitrogen: A Bayesian and machine learning approach. *Agricultural and Forest Meteorology* 311, 108668. <https://doi.org/10.1016/j.agrformet.2021.108668>

- Efron, B., 1990. Six Questions Raised by the Bootstrap (Technical Report No. 139). Stanford, CA: Stanford University, Division of Biostatistics.
- Efron, B., Tibshirani, R., 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science* 1, 54–75.
- Friedrich, S., Groll, A., Ickstadt, K., Kneib, T., Pauly, M., Rahnenführer, J., Friede, T., 2023. Regularization approaches in clinical biostatistics: A review of methods and their applications. *Stat Methods Med Res* 32, 425–440. <https://doi.org/10.1177/09622802221133557>
- Galloway, J.N., Aber, J.D., Erisman, J.W., Seitzinger, S.P., Howarth, R.W., Cowling, E.B., Cosby, B.J., 2003. The Nitrogen Cascade. *BioScience* 53, 341–356. [https://doi.org/10.1641/0006-3568\(2003\)053\[0341:TNC\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2)
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press, Cambridge, Massachusetts.
- Howson, C., Urbach, P., 2005. Scientific Reasoning: The Bayesian Approach, 3rd edition. ed. Open Court, Chicago.
- Jackson, C., 2011. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software* 38, 1–28. <https://doi.org/10.18637/jss.v038.i08>
- Jankowski, K.J., Szatkowski, A., Załuski, D., 2024. The effect of nitrogen management on seed yield and quality in traditional and canola-quality white mustard. *Sci Rep* 14, 26127. <https://doi.org/10.1038/s41598-024-76582-9>
- Jaynes, D.B., 2011. Confidence bands for measured economically optimal nitrogen rates. *Precision Agric* 12, 196–213. <https://doi.org/10.1007/s11119-010-9168-3>
- Johnson, J., Giraud-Carrier, C., 2019. Diversity, accuracy and efficiency in ensemble learning: An unexpected result. *Intelligent Data Analysis* 23, 297–311. <https://doi.org/10.3233/IDA-183934>
- Kuncheva, L.I., Whitaker, C.J., 2003. Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy. *Machine Learning* 51, 181–207. <https://doi.org/10.1023/A:1022859003006>
- Kyveryga, P.M., Blackmer, A.M., Morris, T.F., 2007. Disaggregating Model Bias and Variability when Calculating Economic Optimum Rates of Nitrogen Fertilization for Corn. *Agronomy Journal* 99, 1048–1056. <https://doi.org/10.2134/agronj2006.0339>
- Lobell, D.B., 2007. The cost of uncertainty for nitrogen fertilizer management: A sensitivity analysis. *Field Crops Research* 100, 210–217. <https://doi.org/10.1016/j.fcr.2006.07.007>
- Lyons, S.E., Tang, Z., Booth, J., Ketterings, Q.M., 2019. Nitrogen response models for winter cereals grown for forage. *Journal of Agronomy and Crop Science* 205, 248–261. <https://doi.org/10.1111/jac.12310>
- Mac Nally, R., Duncan, R.P., Thomson, J.R., Yen, J.D.L., 2018. Model selection using information criteria, but is the “best” model any good? *Journal of Applied Ecology* 55, 1441–1444. <https://doi.org/10.1111/1365-2664.13060>

- Matavel, C.E., Meyer-Aurich, A., Piepho, H.-P., 2024. Model-averaging as an accurate approach for ex-post economic optimum nitrogen rate estimation. *Precision Agric* 25, 1324–1339. <https://doi.org/10.1007/s11119-024-10113-4>
- Miguez, F.E., Poffenbarger, H., 2022. How can we estimate optimum fertilizer rates with accuracy and precision? *Agricultural & Environmental Letters* 7, e20075. <https://doi.org/10.1002/ael2.20075>
- Morris, T.F., Murrell, T.S., Beegle, D.B., Camberato, J.J., Ferguson, R.B., Grove, J., Ketterings, Q., Kyveryga, P.M., Laboski, C.A.M., McGrath, J.M., Meisinger, J.J., Melkonian, J., Moebius-Clune, B.N., Nafziger, E.D., Osmond, D., Sawyer, J.E., Scharf, P.C., Smith, W., Spargo, J.T., van Es, H.M., Yang, H., 2018. Strengths and Limitations of Nitrogen Rate Recommendations for Corn and Opportunities for Improvement. *Agronomy Journal* 110, 1–37. <https://doi.org/10.2134/agronj2017.02.0112>
- Ngo, G., Beard, R., Chandra, R., 2022. Evolutionary bagging for ensemble learning. *Neurocomputing* 510, 1–14. <https://doi.org/10.1016/j.neucom.2022.08.055>
- R Core Team, 2020. R: A language and environment for statistical computing.
- Ransom, C.J., Clark, J., Bean, G.M., Bandura, C., Shafer, M.E., Kitchen, N.R., Camberato, J.J., Carter, P.R., Ferguson, R.B., Fernández, F., Franzen, D.W., Laboski, C.A.M., Myers, D.B., Nafziger, E.D., Sawyer, J.E., Shanahan, J., 2021. Data from a public–industry partnership for enhancing corn nitrogen research. *Agronomy Journal* 113, 4429–4436. <https://doi.org/10.1002/agj2.20812>
- Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Sieling, K., Rose, Till, Stahl, Andreas, Reddersen, Björn, and Kage, H., 2023. Differences in N use efficiency, N translocation and N immobilization capacity of their residues of oilseed rape varieties due to N fertilization. *Archives of Agronomy and Soil Science* 69, 3300–3313. <https://doi.org/10.1080/03650340.2023.2228257>
- Sinclair, T.R., Rufty, T.W., 2012. Nitrogen and water resources commonly limit crop yield increases, not necessarily plant genetics. *Global Food Security* 1, 94–98. <https://doi.org/10.1016/j.gfs.2012.07.001>
- Soloff, J.A., Barber, R.F., Willett, R., 2024. Bagging Provides Assumption-free Stability. *Journal of Machine Learning Research* 25, 1–35.
- Ugirimurera, J., Bensen, E.A., Severino, J., Sanyal, J., 2024. Addressing bias in bagging and boosting regression models. *Sci Rep* 14, 18452. <https://doi.org/10.1038/s41598-024-68907-5>
- USDA ERS, 2024. USDA ERS - Documentation and Data Sources [WWW Document]. URL <https://www.ers.usda.gov/data-products/fertilizer-use-and-price/documentation-and-data-sources> (accessed 4.25.24).
- USDA ERS Price Forecast, 2024. USDA ERS - Season-Average Price Forecasts [WWW Document]. URL <https://www.ers.usda.gov/data-products/season-average-price-forecasts> (accessed 4.25.24).

- van Grinsven, H.J.M., Ebanyat, P., Glendining, M., Gu, B., Hijbeek, R., Lam, S.K., Lassaletta, L., Mueller, N.D., Pacheco, F.S., Quemada, M., Bruulsema, T.W., Jacobsen, B.H., ten Berge, H.F.M., 2022. Establishing long-term nitrogen response of global cereals to assess sustainable fertilizer rates. *Nat Food* 3, 122–132. <https://doi.org/10.1038/s43016-021-00447-x>
- Ver Hoef, J.M., 2012. Who Invented the Delta Method? *The American Statistician* 66, 124–127. <https://doi.org/10.1080/00031305.2012.687494>
- Ver Hoef, J.M., Boveng, P.L., 2015. Iterating on a single model is a viable alternative to multimodel inference. *The Journal of Wildlife Management* 79, 719–729. <https://doi.org/10.1002/jwmg.891>
- Wang, G.L., Ye, Y.L., Chen, X.P., Cui, Z.L., 2014. Determining the optimal nitrogen rate for summer maize in China by integrating agronomic, economic, and environmental aspects. *Biogeosciences* 11, 3031–3041. <https://doi.org/10.5194/bg-11-3031-2014>
- Wood, D., Mu, T., Webb, A.M., Reeve, H.W.J., Luján, M., Brown, G., 2023. A Unified Theory of Diversity in Ensemble Learning. *Journal of Machine Learning Research* 24, 1–49.
- Zhang, C., Ma, Y., 2012. Ensemble Learning, in: *Ensemble Machine Learning: Methods and Applications*. Springer, New York, NY, pp. 1–34. https://doi.org/10.1007/978-1-4419-9326-7_1

Appendix A

Supplementary Note 1

A general expression of the Delta Method for estimating the variance of a function of a random variable is given by

$$Var\ g(\mathbf{y}) = \sum_i [g'_i(\mathbf{a})]^2 Var\ y_i + 2 \sum_{i>j} g'_i(\mathbf{a}) g'_j(\mathbf{a}) Cov(y_i, y_j).$$

For the quadratic and quadratic plateau models, let $\mathbf{y} = \hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)^T$, $\mathbf{a} = (E[\hat{\beta}_1], E[\hat{\beta}_2])^T = (\tau_1, \tau_2)^T$, and $\hat{\theta} = g(\hat{\boldsymbol{\beta}}) = \frac{\hat{\beta}_1 - PR}{2\hat{\beta}_2}$. Furthermore, let $\frac{\partial g}{\partial \hat{\beta}_1} = \frac{1}{2\hat{\beta}_2}$, and $\frac{\partial g}{\partial \hat{\beta}_2} = -\frac{2(\hat{\beta}_1 - PR)}{(2\hat{\beta}_2)^2}$. Then the variance of $\hat{\theta} = \widehat{EONR}$ can be approximated by

$$Var\ g(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^2 \left[\frac{\partial g}{\partial \hat{\beta}_i} \Big|_{\hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} \right]^2 Var\ \hat{\beta}_i + 2 \frac{\partial g}{\partial \hat{\beta}_1} \Big|_{\hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} \frac{\partial g}{\partial \hat{\beta}_2} \Big|_{\hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} Cov(\hat{\boldsymbol{\beta}}).$$

For the Mitscherlich model, let $\mathbf{y} = \hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)^T$, $\mathbf{a} = (E[\hat{\beta}_0], E[\hat{\beta}_1], E[\hat{\beta}_2])^T = (\tau_0, \tau_1, \tau_2)^T$, and $\hat{\theta} = g(\hat{\boldsymbol{\beta}}) = -\frac{\ln\left[\frac{-p_f}{p_c(\beta_1 - \beta_0)\beta_2}\right]}{\beta_2}$. In this case we have three partial derivatives as

$$\frac{\partial g}{\partial \hat{\beta}_0} = \frac{p_f(p_c\beta_2)}{[p_c(\beta_1 - \beta_0)\beta_2]^2} \frac{1}{\frac{1}{-p_f[p_c(\beta_1 - \beta_0)\beta_2]}/\beta_2}, \text{ and } \frac{\partial g}{\partial \hat{\beta}_1} = -\left\{ \frac{p_f(p_c\beta_2)}{[p_c(\beta_1 - \beta_0)\beta_2]^2} \frac{1}{\frac{1}{-p_f[p_c(\beta_1 - \beta_0)\beta_2]}/\beta_2} \right\}, \text{ and}$$

$$\frac{\partial g}{\partial \hat{\beta}_2} = \frac{p_f p_c(\beta_1 - \beta_0)}{[p_c(\beta_1 - \beta_0)\beta_2]^2} \frac{1}{\frac{1}{-p_f[p_c(\beta_1 - \beta_0)\beta_2]}/\beta_2} - \frac{\ln(-p_f p_c(\beta_1 - \beta_0)\beta_2)}{\beta_2^2}. \text{ Then the variance of } \hat{\theta} =$$

$\widehat{EONR}$ can be approximated by

$$Var\ g(\hat{\boldsymbol{\beta}}) = \sum_{i=0}^2 \left[\frac{\partial g}{\partial \hat{\beta}_i} \Big|_{\hat{\beta}_0=\tau_0, \hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} \right]^2 Var\ \hat{\beta}_i + 2 \frac{\partial g}{\partial \hat{\beta}_0} \Big|_{\hat{\beta}_0=\tau_0, \hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} \frac{\partial g}{\partial \hat{\beta}_1} \Big|_{\hat{\beta}_0=\tau_0, \hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} \frac{\partial g}{\partial \hat{\beta}_2} \Big|_{\hat{\beta}_0=\tau_0, \hat{\beta}_1=\tau_1, \hat{\beta}_2=\tau_2} Cov(\hat{\boldsymbol{\beta}}).$$

For the linear plateau model, consider that we have the model $y_i = f(x_i, \boldsymbol{\beta}) + \varepsilon_i$, where $\varepsilon \sim n(0, \sigma^2)$. In this case $f(x_i, \boldsymbol{\beta}) = \begin{cases} \beta_0 + \beta_1 x_i & \text{if } x_i < xp \\ \beta_0 + \beta_1 xp & \text{o.w.} \end{cases}$. Then $\boldsymbol{\beta} = (\beta_0, \beta_1, xp)^T$. Then,

since the Gauss-Newton algorithm was implemented,

$$Cov(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{J}^T \mathbf{J})^{-1},$$

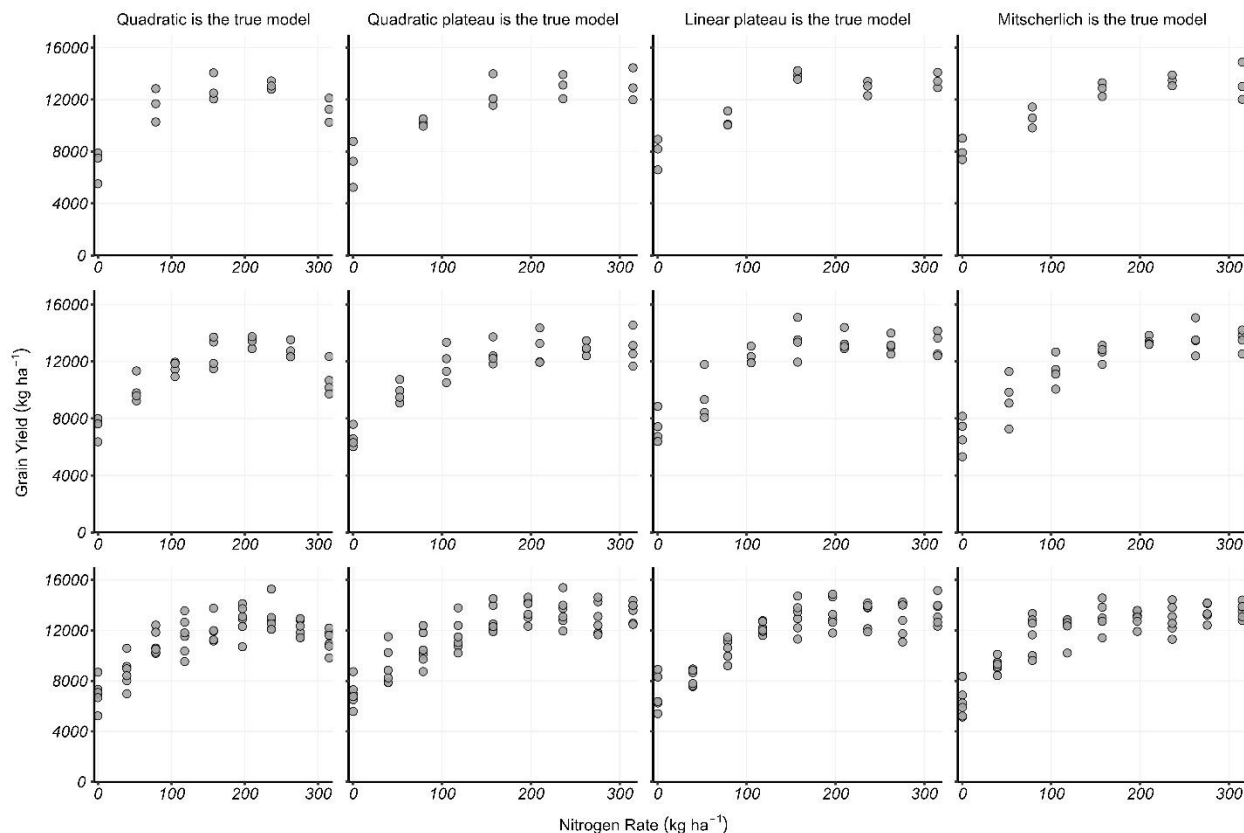
where $\hat{\sigma}^2$ is the mean squared error, $\mathbf{J}$ is the Jacobian matrix of $f(x, \boldsymbol{\beta})$, and $\mathbf{J}^T \mathbf{J}$ is the approximation of the Hessian matrix of the sum of squared residuals (similar to $\mathbf{X}^T \mathbf{X}$ in linear models). Therefore, in this case variance of $\hat{\theta} = \widehat{EONR} = \widehat{xp}$ is the third element of the diagonal of the matrix $Cov(\hat{\boldsymbol{\beta}})$.

In all cases, the 95% confidence intervals of $\hat{\theta}$ were approximated as

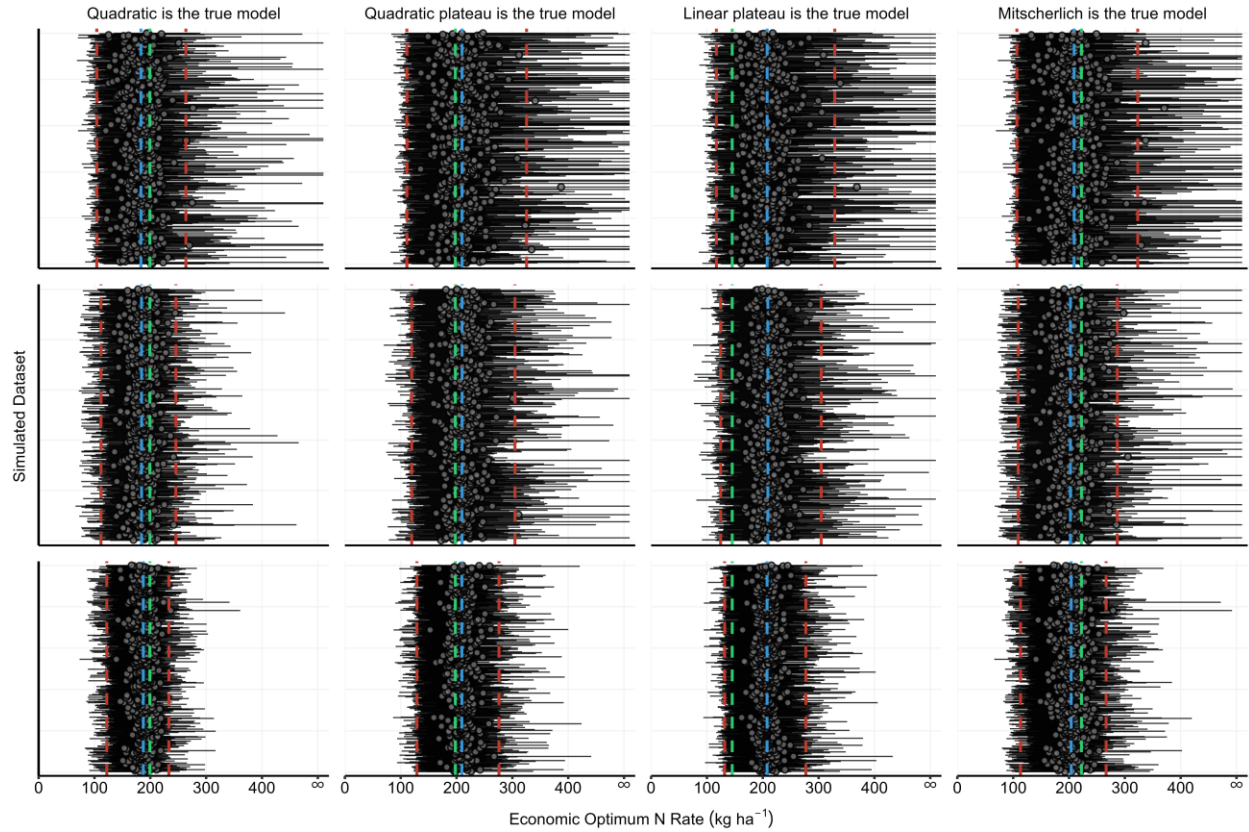
$$CI_{\theta} = \left(\hat{\theta} - 1.96 \sqrt{\widehat{Var}[\hat{\theta}]} < \theta < \hat{\theta} + 1.96 \sqrt{\widehat{Var}[\hat{\theta}]} \right).$$

Appendix B

Supplementary Figures



Supp. Figure 1. Representative examples of simulated data for the relationship between maize grain yield and nitrogen rate. Each plot represents one of the twelve scenarios created by the combination of the data generating process (true model) and data quality. The model generating the data is in the top of each column. The first row of plots indicates low data quality, the middle row is for intermediate data quality, and the bottom row indicate high data quality.



Supp. Figure 2. 95% confidence intervals of the EONR estimation obtained by using bagging for the 500 simulated data sets. The red vertical lines indicate the median of the lower and upper bounds of the confidence intervals across the 500 datasets. The blue vertical line indicates the estimated EONR. The green vertical line represents the true value of the EONR. Each plot represents one of the twelve scenarios created by the combination of the data generating process (true model) and data quality. The model generating the data is in the top of each column. The first row of plots indicates low data quality, the middle row is for intermediate data quality, and the bottom row indicate high data quality. When the upper bound of the confidence interval was greater than 500 kg ha⁻¹ the infinity (∞) symbol was set, indicating non-agronomic sense for those values.

Appendix C

R Code for Bagging

```
## Packages
library(tidyverse)

## Define the functions for fitting non linear models

# Quadratic Plateau Model
quad_p <- function(x,b0,b1,b2){

  xp = b1/(2*b2)
  x_2 = x^2
  xp_2 = xp^2

  if_else(x < xp,
          b0 + (b1*x) - (b2*x_2),
          b0 + (b1*xp) - (b2*xp_2)
  )

}

# Linear Plateau Model
lin_p <- function(x,b0,b1,EONR){

  if_else(x < EONR,
          b0 + (b1*x),
          b0 + (b1*EONR)
  )

}

# Mitscherlich Model
mits <- function(x,b0,b1,b2){

  b0+(b1-b0)*exp(-b2*x)

}

## Generate bootstrap datasets

df_to_boot <- data.frame(Nrate_kg.ha = df$N_rate,
                        GY_kg.ha = df_ss$Grain_Yield)

set.seed(59)
K <- 5000 # Number of bootstraps

# Perform bootstrapping on a single data frame
df_boot <- replicate(K, df_to_boot[sample(nrow(df_to_boot), replace = TRUE),
],
                    simplify = FALSE)
```

```

for (i in 1:K) {

  df_boot[[i]] <- df_boot[[i]] %>%
    mutate(rowname = NULL,
           boot_id = i,
           .before = Nrate_kg.ha) %>%
    as_tibble()

}

## Apply bagging

# Set a seed for reproducibility
set.seed(497)

# Define the fertilizer and corn prices, and calculate the price ratio
pN = 0.789
pC = 0.154
PR = pN/pC

# Create a list to store the model outputs
bagg_models <- list()

for (i in 1:K) {

  # Sample the model to be fitted in the lth data set
  mtofit <- sample(c("QD", "QP", "LP", "MIT"),
                  size = 1,
                  prob = c(0.25, 0.25, 0.25, 0.25)
  )

  # Initialize an empty model object
  fitM <- NULL

  # Fitting the selected model catching errors
  fitM <- tryCatch(
    {
      if (mtofit == "QD") {
        # Fit the quadratic model
        lm(GY_kg.ha ~ Nrate_kg.ha + I(Nrate_kg.ha^2),
           data = df_boot[[i]])

      } else if (mtofit == "QP") {
        # Fit the quadratic plateau model
        nls(GY_kg.ha ~ quad_p(x = Nrate_kg.ha, b0, b1, b2),
            data = df_boot[[i]],
            start = list(b0 = 7500, b1 = 55, b2 = 0.12),
            control = list(maxiter = 2000))

      } else if (mtofit == "LP") {
        # Fit the linear plateau model
        nls(GY_kg.ha ~ lin_p(x = Nrate_kg.ha, b0, b1, EONR),
            data = df_boot[[i]],
            start = list(b0 = 7500, b1 = 40, EONR = 160),
            control = list(maxiter = 2000))

      }
    }
  )
}

```

```

    } else if (mtofit == "MIT") {
      # Fit the Mitscherlich model
      nls(GY_kg.ha ~ mits(x = Nrate_kg.ha, b0, b1, b2),
        data = df_boot[[i]],
        start = list(b0 = 14000, b1 = 7000, b2 = 0.01),
        control = list(maxiter = 2000))
    }
  },
  error = function(e) NA) # Return NA if an error occurs

# Store results only if the model fit was successful
bagg_models[[i]] <- tryCatch(
  {
    tibble(
      model = mtofit,
      b0 = as.numeric(coef(fitM)[1]),
      b1 = as.numeric(coef(fitM)[2]),
      b2 = case_when(mtofit == "LP" ~ NA_real_,
                     TRUE ~ as.numeric(coef(fitM)[3])),
      EONR = case_when(mtofit == "LP" ~ as.numeric(coef(fitM)[3]),
                       mtofit == "MIT" ~ -log(-pN / (pC * (b1 - b0) * b2))
                       / b2,
                       TRUE ~ (b1 - PR) / (2 * abs(b2)))
    )
  },
  error = function(e) NA)
}

# Remove NA elements from the list
bagg_models_NONA <- bagg_models[!sapply(bagg_models, function(x) is.null(x)
|| identical(x, NA) || !inherits(x, "data.frame"))]

# Bind the model outputs
df_bagg <- bind_rows(bagg_models_NONA)

```