agronomy

MDPI

Article

Advancing toward Personalized and Precise Phosphorus Prescription Models for Soybean (*Glycine max* (L.) Merr.) through Machine Learning

Floyd Muyembe Chipatela [1,*], Lotfi Khiari [1,2], Hamza Jouichat [2], Ismail Kouera [1] and Mahmoud Ismail [3]

[1] Center of Excellence for Soil and Fertilizer Research in Africa, College of Agriculture and Environmental Sciences, Mohammed VI Polytechnic University, Benguerir 43150, Morocco; lotfi.khiari@um6p.ma (L.K.); ismail.kouera@um6p.ma (I.K.)
[2] Department of Soil Science and Agrifood Engineering, Laval University, Quebec, QC G1V 0A6, Canada; hamza.jouichat.1@ulaval.ca
[3] EMINES School of Industrial Management, Mohammed VI Polytechnic University, Benguerir 43150, Morocco; mahmoud.ismail@emines.um6p.ma
* Correspondence: chipatela.floyd@um6p.ma; Tel.: +212-709-871-379 or +260-963-493-140

Abstract: The traditional approach of prescribing phosphate fertilizer solely based on soil test P (STP) has faced criticism from scientists and agriculturists pushing farmers to seek phosphate fertilization models that incorporate additional factors. By embracing integrated approaches, farmers can receive more precise recommendations that align with their specific conditions and fertilization techniques. This study aimed to utilize artificial intelligence prediction to replicate soybean response curves to fertilizer by integrating edaphic and climatic factors. Literature data on soybean response to P fertilization were collected, and the Random Forest (RF) algorithm was applied to predict response curves. The predictions utilized seven predictors: P dose, STP, soil pH, texture, % OM, precipitation, and P application methods. These predictions were compared to the traditional STP-based approach. The STP-based P prescription models exhibited extremely low robustness values (R^2) of 1.53% and 0.88% for the P_{Bray-1} and P_{Olsen} diagnostic systems, respectively. In contrast, implementing the RF algorithm allowed for more accurate prediction of yield gains at various P doses, achieving robustness values of 87.4% for the training set and 60.9% for the testing set. The prediction errors remained below 10% throughout the analysis. Implementing artificial intelligence modeling enabled the study to achieve precise predictions of the optimal P dose and customized fertilization recommendations tailored to farmers' specific soil conditions, climate, and individual fertilization practices.

Keywords: optimal field-specific rate; predicting response curves; prescription of P; unified phosphorus fertility classification system

check for updates

Citation: Chipatela, F.M.; Khiari, L.; Jouichat, H.; Kouera, I.; Ismail, M. Advancing toward Personalized and Precise Phosphorus Prescription Models for Soybean (*Glycine max* (L.) Merr.) through Machine Learning. *Agronomy* **2024**, *14*, 477. https://doi.org/10.3390/agronomy14030477

Academic Editor: Roberto Marani

Received: 1 January 2024
Revised: 18 January 2024
Accepted: 23 January 2024
Published: 27 February 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Soybean (*Glycine max* (L.) Merrill) is the most common legume seed in the world. It is an important agronomic and economical crop cultivated worldwide for its high protein and oil content for human nutrition, poultry feed, livestock, and aquaculture industries [1]. Therefore, countries with less fertile land should increase their fertilizer inputs to meet soybean demand. Thus, optimal fertilizer management for specific regional conditions is essential for soybean and food security.

Phosphorus (P) is often one of the most growth-limiting nutrients for crop production in many soils. An adequate P fertilization is necessary to attain optimum yields [2,3]. Phosphorus uptake by soybean is roughly 5.5 g P per kg of soybean, equivalent to fertilizer P uptake of 9–10 kg P ha^{-1} for a target yield of 1700 kg ha^{-1} [4]. However, since low-fertility soils are generally high P-fixing in nature, inputs are from three to five times this soybean P requirement. Although adding P fertilizer has been shown to increase crop yields, the

response of soybean to this nutrient still varies from place to place and from study to study. Indeed, some studies report little or no yield response to P fertilization in low P testing soils, and other studies show high response in high P testing soils [5–8].

Traditionally, soil test phosphorus (STP) has been used as the single guideline for prescribing phosphorus rates. This traditional concept of P recommendation based on STP is the most widely applied for all crops and regions worldwide [2,9,10]. Several STP methods estimate available phosphorus in the soil, with Bray-1, Olsen, and Mehlich-3 among the common ones. However, the P extraction methods do not apply to all soil types. For example, Bray-1 and Mehlich-3 extractants are designed to extract available P from non-calcareous soils [10–12], whereas the Olsen extraction method is meant for calcareous soils [13]. Regardless of the P extraction method, the "traditional" fertilizer recommendation process involves correlation, calibration, and optimization. The correlation comes down to the choice of the STP method. At the same time, calibration assigns a critical value to the selected STP to distinguish between responsive or unresponsive soils to P fertilization [14]. The critical STP value varies with factors such as the soil properties, soil depth, crop species, climate, and choice of partition model [14,15].

The most widely used model for determining the critical STP is the Cate-Nelson. This model relates STP values to relative yields or absolute yield increases [16,17]. The Cate-Nelson model is effective in statistically discriminating between fields with a high response to fertilization from those that respond less [14]. Once we have obtained this critical STP by calibration, we use it to construct phosphorus fertility classes. The final step is optimization. Thus, we deduce optimal amounts of P from the phosphorus response curves site by site. These optimal doses are then grouped into each fertility class to derive a mean or median per class. Traditionally, we end up with a simplified fertilization grid at an average optimal dose per fertility class. The influence of P input on crop yield is often discussed based on the shape of a response curve. The discussion usually focuses on the pattern of point distributions, whether they follow a plateau or a linear or curvilinear trend. According to [18], the response curves to P fertilization can vary depending on the fertilization strategy, tillage system, and varieties used. Some past studies have used data from experimental trials to develop P fertilization models for potatoes, maize, and soybean by following the steps of correlation, calibration, optimization, and discretization [19,20].

The response curve is undoubtedly the core component of the fertilizer recommendation system and the part that has attracted much criticism. These criticisms revolve around the choice of mathematical models that best represent the response function and thus add precision in choosing the economically optimal rates. Linear-plateau, quadratic, cubic, and Mitscherlich–Bray equations are generally used to estimate fertilizer response curves [20,21]. There are no established rules for the choice of the fitting equation, but the correlation coefficient (R^2) is commonly used to judge the quality of the fit. These equations are then transformed into a first derivative equated to the productivity limit to deduce an economically optimal P application rate per experimental site. Waugh et al. [22] showed the variation in the selection of the economically optimal rate by fitting several of these equations to the same experimental data. They reported that the quadratic equations commonly used to describe response curves tend to overestimate the maximum fertilizer rates.

The biggest disadvantage of the traditional fertilizer grid approach is that it reduces the response curve to a single piece of information. This economically optimal P rate is then amalgamated with other rates within the same fertility class to obtain an average or median. Once P rates are determined, the response curves are no longer available for updating as fertilizer and crop prices change despite the high cost of scientific experimentation to obtain them. This double dilution of specific information takes us away from the right economically optimal P rate, especially since it is well known that soybean response to P fertilizer varies from place to place [6,23,24]. Therefore, with the present trend of ever-changing prices of fertilizers and crop production, there is a need to create a system that predicts response curves and infers optimum P rates using real-time data.

A traditional recommendation model for a given element is based solely on the soil content of that element, which implies that yield response depends on that element alone and that any other fertility indicators or management practices do not have an influence [25]. However, crop response to fertilizer varies depending on soil, climate, and crop type factors [2,3,26]. Machine Learning (ML) is one such advancement that allows the incorporation of factors other than STP in predicting the optimal fertilizer recommendation rates. ML learns the algorithms based on supervised, unsupervised, or semi-supervised learning [27] and can help to improve the accuracy of any given recommendation model. Coulibali et al. [28] applied ML in predicting nitrogen, phosphorus, and potassium requirements for high potato tuber yield using climatic, edaphic, and field management variables. Similarly, Abera et al. [29] applied the Random Forest model to predict wheat yield response to fertilizer by incorporating climatic, topographic, and edaphic factors. Manoj et al. [30] designed a fertilizer recommendation model using the SVM model in Machine Learning, describing the best suitable crop and fertilizer to be applied depending on soil and weather conditions.

Fertilizer recommendation models are specific to crops. Our literature search did not present any past studies that have used Machine Learning to prescribe P fertilizer application rates for soybean. We see that before 2020, recommendation models were based on the classical approach [19,20], while studies in the last 2 years [28–30] have bypassed the classical approach to introduce ML algorithms to increase the robustness of fertilizer application rate prediction. As such, our study will explore both directions and show that fertilizer recommendation can be enhanced by incorporating edaphic, climatic, and management factors in the ML environment. Our study will show how to move from the traditional (average rate approach) to the more site-specific ML approach for soybean cultivation.

2. Materials and Methods

This study consists of building a database from previously published research. These data will be used to establish fertilizer recommendation models according to the classical one-factor approach based only on STP. The same data will also be used to develop a Machine Learning algorithm to predict fertilizer recommendations based on several factors (edaphic, climatic, and field management practices). Figure 1 shows the schematic representation of the methodology adopted for the study.

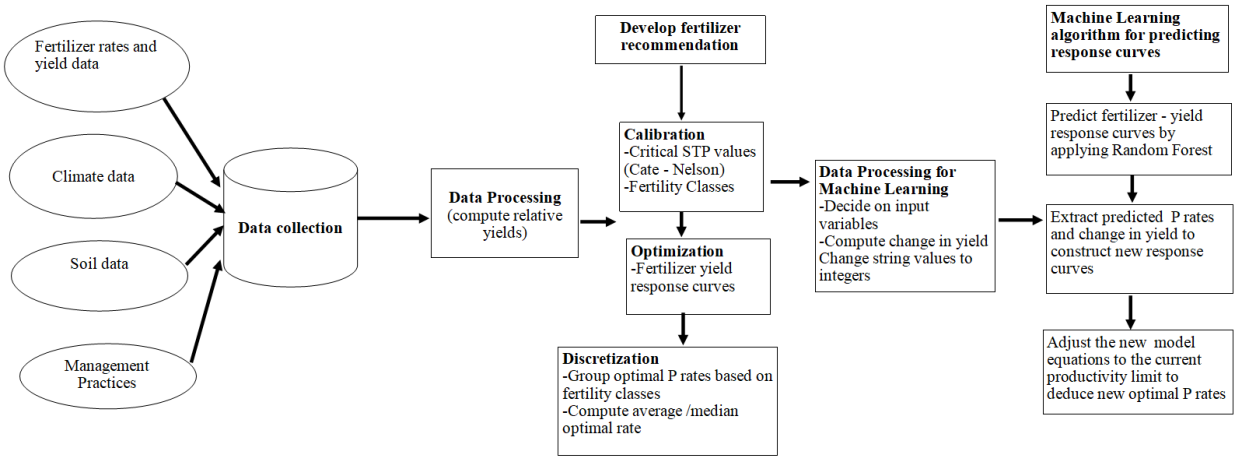

Figure 1. Schematic representation of the methodology.

2.1. The Database

The study assembled a database comprising data about soybean yield response to fertilizer P application for the various studies conducted around the globe. The database was built and managed in Microsoft Excel. These data were sourced from peer-reviewed

and published literature, as summarized in Table 1. As such, a rich diversity in data sources was achieved and encompassed various soil and climate conditions with experiments conducted between 1968 and 2018. Three key criteria were set for data to be included in the database: (i) data from field experiments with a control treatment (non-fertilized) and a fertilized treatment or more, (ii) details of fertilizer management (rate and method of application), and (iii) a soil test phosphorus (STP) of any commonly used P extraction method. Web-plot digitizer (desk-top version) was used to extract data from graphs, charts, and other uneditable sources. The key data variables extracted from the articles included soil characteristics (available soil phosphorus, soil organic matter, soil pH, soil texture, and soil type), climatic factors (precipitation and temperature), method of fertilizer application (broadcasting, banding, or drilling), and the cultivar of soybean used in the experiment. In this study, soybean yield is reported in kilogram per hectare (kg ha^{-1}), fertilizer application rates as kg P ha^{-1}, STP in milligrams per kilogram (mg kg^{-1}) soil organic carbon in percent (%), precipitation in millimeters (mm), and atmospheric temperature in degrees Celsius (°C). As such, data from articles that reported these parameters in other units had to be converted accordingly. Additional information recorded included the country and location where the experiment was conducted, including GPS coordinates and the year in which the experiment was conducted. Soil texture was categorized into three textural groups, including G1 for fine-textured soils (heavy clay, clay, silty clay, silty clay loam, clay loam, sandy clay, sandy clay loam), G2 for medium-textured soils (loam, silt loam, silt), and G3 for coarse-textured soils (sand, loamy sand, sandy loam). Most of the experiments in the articles sourced for data used either Bray-P1 (P_{Bray-1}) or Olsen P (P_{olsen}) extraction methods. Thus, we decided to use only the data from articles that reported to have used either P_{Bray-1} or P_{Olsen} extraction methods in the study. The literature search managed to source data from 39 peer-reviewed articles. The 39 articles gave a total of 219 global observation points. From the 219 observations, 90 and 129 were from articles that employed P_{Olsen} and P_{Bray-1} extraction methods for STP, respectively. These observations came from a total of 69 separate experiments (Table 1).

Table 1. References from the soybean phosphate fertilization trial database collected from several regions with soil P diagnostic systems based on P_{Bray-1} and P_{Olsen} methods.

Source	Points	Exps	STP	Country
[31]	4	1	Bray-1	USA
[32]	30	6	Bray-1/Olsen	USA
[33]	4	1	Bray-1	Argentina
[34]	4	1	Bray-1	Brazil
[35]	5	1	Bray-1	Nigeria
[36]	3	1	Bray-1	South Africa
[37]	8	2	Bray-1	Nigeria
[38]	4	1	Bray-1	India
[39]	4	1	Bray-1	India
[40]	4	1	Bray-1	Ethiopia
[1]	2	1	Bray-1	Ghana
[41]	2	1	Bray-1	Argentina
[42]	2	1	Bray-1	Ghana
[43]	4	2	Bray-1	Ghana
[44]	2	1	Bray-1	USA
[45]	6	3	Bray-1	Ghana
[46]	2	1	Bray-1	Ghana
[47]	4	1	Bray-1	Nigeria
[48]	8	2	Bray-1	Benin
[49]	4	1	Bray-1	Bangladesh
	106	**30**		

Table 1. *Cont.*

Source	Points	Exps	STP	Country
[50]	5	1	Bray-1	Nigeria
[51]	24	12	Bray-1	Nigeria
[52]	5	1	Bray-1	India
[53]	4	2	Bray-1	Benin
[54]	4	1	Olsen	Bangladesh
[55]	5	1	Olsen	Pakistan
[56]	4	1	Olsen	Bangladesh
[57]	4	1	Olsen	Bangladesh
[58]	4	1	Olsen	Pakistan
[59]	5	1	Olsen	India
[60]	4	1	Olsen	Nigeria
[61]	4	1	Olsen	Ethiopia
[62]	4	1	Olsen	India
[63]	3	1	Olsen	Pakistan
[64]	3	1	Olsen	Pakistan
[65]	10	2	Olsen	USA
[66]	3	1	Olsen	Nigeria
[67]	16	8	Olsen	Nigeria
[68]	2	1	Olsen	China
	113	**39**		

Exps: number of experiments reported by the article; **STP:** extraction method used for plant available phosphorus; **Points**: number of points, fertilizer rates or treatments observed in a particular article. The various soil types reported in the articles sourced for data included Oxisols, Alfisols, Acrisols, Inceptisols, Utisols, Nitisols, Luvisols, Vertisols, Acrisols, Ferralsols, Entisols, Ultisols, Alfisols, Gleysols, Lixisols, Cambisols, Lithosols, Mollisols, and Inceptisols. Among the soybean cultivars used in the various experiments comprising the database included BARI-6, NARC-1, PK 416, Williams, TGX varieties, Nandou-12, Hi-Soy2846, Sambaiba, M281, JS-7105, JS-335, Dhidhessa, Jenguma, Lee, and Samsoy 2.

2.2. Determination of Critical STP Concentration and Delineation of Soil Fertility Classes

The critical STP concentration is the plant available soil P value below which there is a high likelihood of crop yield response to P fertilization and vice versa [15]. The Cate-Nelson graphical technique [16] in R statistical software (version 3.6.2) was used to determine the critical STP levels for the $P_{Bray\text{-}1}$ and P_{Olsen} STP diagnostics. The critical STP values were determined from the relationship between relative yields (RY %) and STP values. Relative yield was calculated as the ratio of yield in the control treatment (Yo) to the maximum yield (Ymax) observed in the other P-involved treatments. The Cate-Nelson procedure sets all the relative yield values against the STP values on a scatter diagram upon which vertical and horizontal lines are superimposed to maximize the number of points in the first and third quadrants [17,25,69]. The performance criteria of the Cate-Nelson partitioning models were calculated using the number of points in the first, second, third, and fourth quadrants, which corresponds to the false-positive (FP), true-negative (TN), false-negative (FN), and true-positive (TP) quadrants. Five parameters explain the performance of the Cate-Nelson model: (i) the robustness, $R^2 = \frac{TP+TN}{TP+TN+FP+FN} \times 100$, refers to the probability of making a correct diagnosis for soil phosphorus; (ii) the specificity, $\frac{TN}{TN+FP} \times 100$, represents the probability of making the correct decision to not fertilize with respect to all observations within the yield stability zone; (iii) the sensitivity, $\frac{TP}{TP+FN} \times 100$, represents the probability of making a good decision to fertilize with respect to all observations with low relative yields; (iv) the positive predictive value, PPV = $\frac{TP}{TP+FP} \times 100$, which is the probability of a

positive response of yield to phosphorus fertilization when the STP content is less than the critical agronomic threshold; and (v) the negative predictive value, NPV = $\frac{TN}{TN+FN} \times 100$, relates to the probability that crops will not respond to phosphorus fertilization when the STP content exceeds the critical agronomical threshold.

The critical STP values enabled us to delineate seven fertility classes for the $P_{Bray\text{-}1}$ and P_{Olsen} STP diagnostics. The critical STPs were used as starting values for constructing the fertility classes following the modified fertility index of Cope and Rouse [15]. This approach allocates a fertility index of 100 to the critical STP concentration and assigns two fertility classes below it, with fertility indexes ranging from 0 to 50% and from 50 to 100%. Additional four fertility classes are placed above the critical STP value and assigned fertility index intervals of 100–150%, 150–200%, 200–250%, 250–400%, and >400%. The seven soil fertility classes include Very Low (VL), Low (L), Medium (M), Medium High (MH), High (H), Very High (VH), and Extremely High (EH).

2.3. Determination of P Fertilizer Recommendations Using the Classical (STP-Based) Approach

We plotted absolute yield as a function of fertilizer rates to determine optimal P rates for each experimental site. This procedure used only experiments having more than two fertilizer application rates to fit the response curves. We fitted four response curve shapes, namely, (i) quadratic, when the slope of the curve (yield/rate) is initially high and then progressively decreases until it cancels out when yield peaks and then becomes negative at very high P rates; (ii) cubic curve, shaped like a sigmoid that starts with a convex portion showing resistance of soybean yield response to low P rates due to P sorption and immobilization, followed by a concave portion where the yield increase becomes less than proportional to P rates. This model is used in situations of soils that are highly P-fixing or considered phosphorus sinks; (iii) linear, where soybean yield is proportional to the P rates; and (iv) linear-plateau, where there is an initial proportional soybean yield response to P rates, and a sadden and clear non-response or plateau. This process adopted the method of Khiari et al. [19], which assumes that the first derivative of the quadratic or cubic equations from each response curve equals the productivity limit (PL), as summarized by Equation (2). The productivity limit was computed as the ratio of the cost of one unit (kg) of fertilizer P (Triple Super-phosphate—TSP) to the price of one unit (kg) of soybeans on the market (Equation (1)).

$$PL = \frac{\text{Price P}\left(\frac{\$}{kg}\right)}{\text{Price Soybean}\left(\frac{\$}{kg}\right)} \quad (1)$$

And therefore:

$$\frac{\partial \text{ Yield}\left(\text{kg soybean ha}^{-1}\right)}{\partial \text{ Rate}\left(\text{kg P ha}^{-1}\right)} = PL \quad (2)$$

The study used the average PL computed for the 10 years from July 2012 to July 2022 [70]. The value used as PL for this study is 4.11 (i.e., taking USD 1.93 and USD 0.47 as the price per kg of P and soybean, respectively). In the case of linear plateau response curves, the optimal P rate was obtained at the intersection between the linear response and yield plateau. For linear models, the highest P rate applied was taken as the optimal P rate if the slope of the linear equation is greater than the productivity limit or zero if the opposite is true [20]. For each soil fertility class (i.e., VL, L, ML, MH, H, VH, and EH), optimum P rates across experimental sites were ranked in ascending order. A median P rate was selected as the optimal P fertilizer recommendation rate. We also presented the raw recommendation models showing the scatterplot of the optimal P rates against the corresponding STPs.

2.4. Point-by-Point Reconstruction of P-Fertilizer Response Curve Using Machine Learning

For the same field with the same predictors of OM, pH, precipitation level, textural group, and phosphate fertility class, the application of ML allows assigning to each P dose

a value of yield gain ΔY of soybeans. We computed the yield gain to fertilizer P for each observation, which was then used in the machine learning process. The yield gain (ΔY) was calculated by subtraction of the yield obtained from the control treatment (Y0) from the yield reported for each P fertilized treatment (Y) as summarized by Equation (3).

$$\Delta Y\left(\frac{kg}{ha}\right) = Y\left(\frac{kg}{ha}\right) - Y0\left(\frac{kg}{ha}\right) \quad (3)$$

The justification for using the yield gain instead of the raw yield in the ML is that the ΔY eliminates the influence of other crop management factors and only accounts for the change in yield due to fertilizer P application. To build a single database for Machine Learning, we assigned each STP value to one of seven fertility classes. This way, we combined the two diagnostic systems, $P_{Bray\text{-}1}$ and $P_{Olsen,}$ into the same ML algorithm with 219 inputs from 69 experiments. We used the Random Forest (RF) to model ΔY ($Y_{fertilized\ plot} - Y_{unfertilized\ plot}$), as it constructs a nonlinear relationship between the yield gain and the predictor variables. Since RF is designed to avoid overfitting and its training results are comparable to testing, we selected it to predict ΔY. We used Python version 3.9.12 [71] to apply the Random Forest regressor by setting the default hyperparameters in the Scikit-Learn package version 1.1.1 [72]. In another aspect, the RF has a relatively robust performance in capturing collinearity among predictor variables and noisy covariate data, in addition to its comparatively better performance than other ML models [29]. The RF model is also among the most used in literature. The RF regressor allowed us to predict the ΔY from observed yield based on P rates and the six other predicting variables, including soil pH, OM, STP-fertility classes, precipitation, texture, and fertilizer application methods. The data were preprocessed to ensure compatibility with the language requirements of the ML environment. Data preprocessing began in Microsoft Excel, where categorical features such as soil fertility classes, soil texture groups, and the method of fertilizer application were preprocessed into numerical features. Data preprocessing continued in Python (Version 3.11) and was done to identify missing values and assign arbitrary ones. For the present study, the database was partitioned into 70% for training and 30% for testing and model accuracy assessment, according to Coulibali et al. [28].

The model building followed a stepwise procedure starting from all potential variables that can explain yield gain, then dropping those variables that do not show variability and do not improve model performance. For both the training and testing scenarios, the predictive power of the RF model was based on the robustness (R^2), slope (m), intercept, and root-mean-square error (RMSE) between the predicted and measured ΔY. The prediction performs better the closer the R^2 and the slope of the line between the predicted and measured values of ΔY are to 1, and its intercept and RMSE are to 0. The target is usually for an R^2 value greater than 0.5 [29]. The RMSE is the square root of the average squared differences between predictions and observations. The RMSE indicates prediction errors in the units of the variable of interest with a zero value, implying a perfect fit. The significance of each predictor variable and its partial dependency in predicting ΔY was assessed using Scikit-Learn [72].

Additionally, the importance of the predictor variables for the training and testing datasets was independently assessed with the permutation feature importance model inspection technique. The partial dependence of ΔY on each of the predictor variables was also calculated using the Scikit-Learn package [29,30,72]. The response curves for the predicted gain yield for each experimental site were also derived using the Scikit-Lean package. Once we have adequately controlled the prediction of the ΔY yield gain by RF, we rebuilt point by point the response curve of the yield gain (ΔY) as a function of the different P rates. We ran the RF code as often as the number of points we chose for the response curve, starting with the control without P, then increasing rates in kg of P applied per ha. We then fitted the linear, linear-plateau, quadratic, or cubic response models depending on the shape of the points predicted by RF. We then deduced the optimal rate by applying the limit productivity principle (Equation (2)).

3. Results

3.1. Descriptive Statistics for Key Variables of Interest

In this database, it was observed that soybeans grow in a wide range of mineral soils with OM ranging from 0.2 to 9.8%, pH ranging from very acidic (4.7) to frankly alkaline (8.0), and precipitation conditions ranging from arid to very wet, with rainfall totals ranging from 372 to 2249 mm per year. The link between soybean yields and rainfall is obvious. However, the distribution of precipitation in the soybean growing season can be challenging. Even if we are in a very wet region, the arrival of a dry period of a few weeks during the critical stages of soybean development is enough to cause significant yield reductions. Unfortunately, the data available does not allow us to work on a fine scale in terms of time. The soybean yields recorded in the database of our study ranged from 213 kg ha^{-1} to 7300 kg ha^{-1}, with a mean yield of 1912 kg ha^{-1}. A high variation in available P was observed in the $P_{Bray\text{-}1}$ and P_{Olsen} STP values recorded in the database for the study, with high standard deviations of about 8.69 mg kg^{-1} and 8.55 mg kg^{-1}, respectively. Table 2 summarizes some of the descriptive statistics of selected variables that formed the database for the study.

Table 2. Main edaphic and climatic properties of the 69 experimental sites collected from scientific papers.

	Mean	Median	Standard Dev	Minimum	Maximum
pH_{water}	6.2	6.1	0.8	4.7	8.0
Organic matter (%)	1.4	0.9	1.7	0.2	9.8
Available $P_{Bray\text{-}1}$ (mg kg^{-1})	10.8	7.1	8.7	0.5	32.2
Available P_{Olsen} (mg kg^{-1})	10.0	6.2	8.6	2.3	27.3
Yield (kg ha^{-1})	1912	1676	936	213	7300
Rainfall (mm)	1094	1092	472	372	2249
Average Temperature (°C)	24.1	24.1	5.6	14.5	33.6

3.2. Cate-Nelson Partitioning and Soil Fertility Classification

We presented the Cate-Nelson two-group partitioning procedure in Figures 2 and 3. This partitioning resulted in two critical STP values of 7.5 mg kg^{-1} for $P_{Bray\text{-}1}$ (Figure 2) and 8.4 mg kg^{-1} for P_{Olsen} (Figure 3). As proposed by Nelson and Anderson [17], we separated plant responses into a group of sites with a strong response to P fertilizer (soils of low and very low phosphate fertility) and another with a weak response (soils of medium and high fertility). This method gives results similar to an analysis of variance that maximizes the sum of squares between these two groups (Figures 2c and 3c). The critical STP values correspond to 90% and 83% relative yields. Figures 2 and 3 show that the high probabilities of economic response for soybean to phosphorus fertilization were in the TP quadrants of low phosphorus contents, i.e., STP less than 7.5 mg $P_{Bray\text{-}1}$ kg^{-1} and 8.4 mg P_{Olsen} kg^{-1}, which correspond to the relative yield range of 20–90% for $P_{Bray\text{-}1}$ and 55–83% for P_{Olsen}. By minimizing the number of points in the error quadrants: FN (false negative: lower right quadrant) + FP (false positive: upper left quadrant) in Figures 2b and 3b, we obtained the limits of 90% and 83% of critical relative yields. Beyond these critical values, we observed the stability of the yields and a weak response to phosphate fertilization. In the TN quadrant, relative yield values reached stability between 90% and 100 (Figure 2a) and between 83% and 100% (Figure 3a). Error quadrants FP corresponded to high relative yields even though they pertained to regions encompassing low phosphorus fertility classes. In contrast, FN quadrants corresponded to low relative yield while falling in areas encompassing high-phosphorus fertility classes (Figures 2b and 3b).

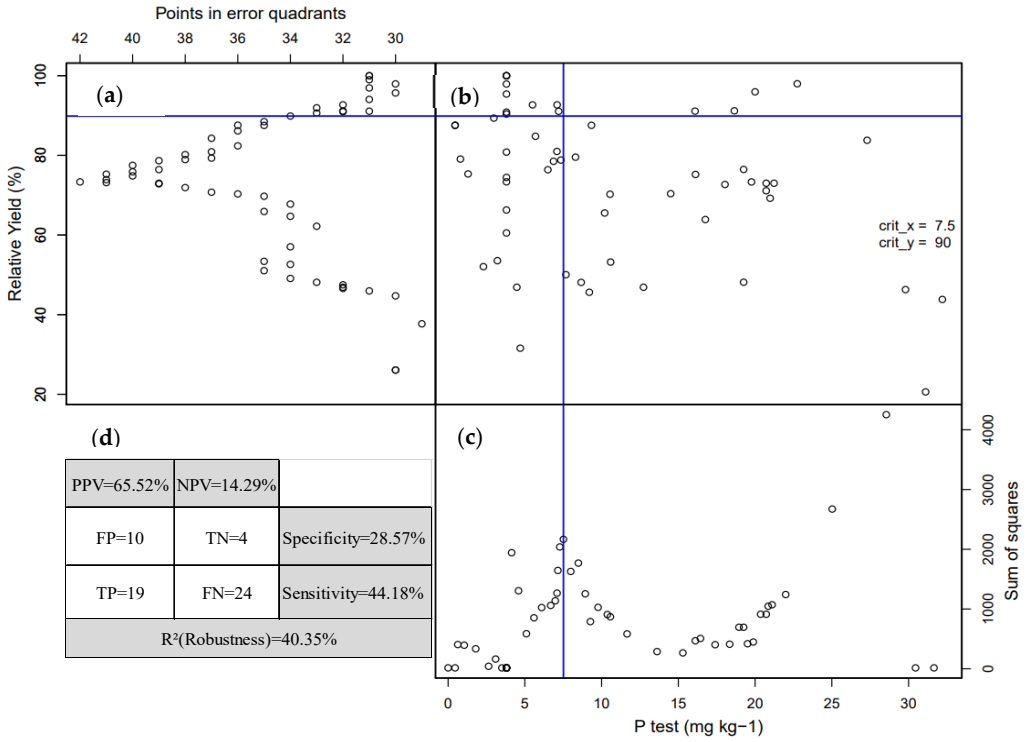

Figure 2. Cate-Nelson partitioning model for P_{Bray-1} showing: (**a**) the number of points outside the model used in determining the critical relative yield (crit_y); (**b**) identified critical levels (crit_x and crit_y); (**c**) the sum of squares used in determining the critical *p* value (crit_x); and (**d**) model summary table (lower left): false positive (FP), true negative (TN), false negative (FN), and true positive (TP)—number of points in the different quadrants; model performance indicators: robustness (R^2), sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV).

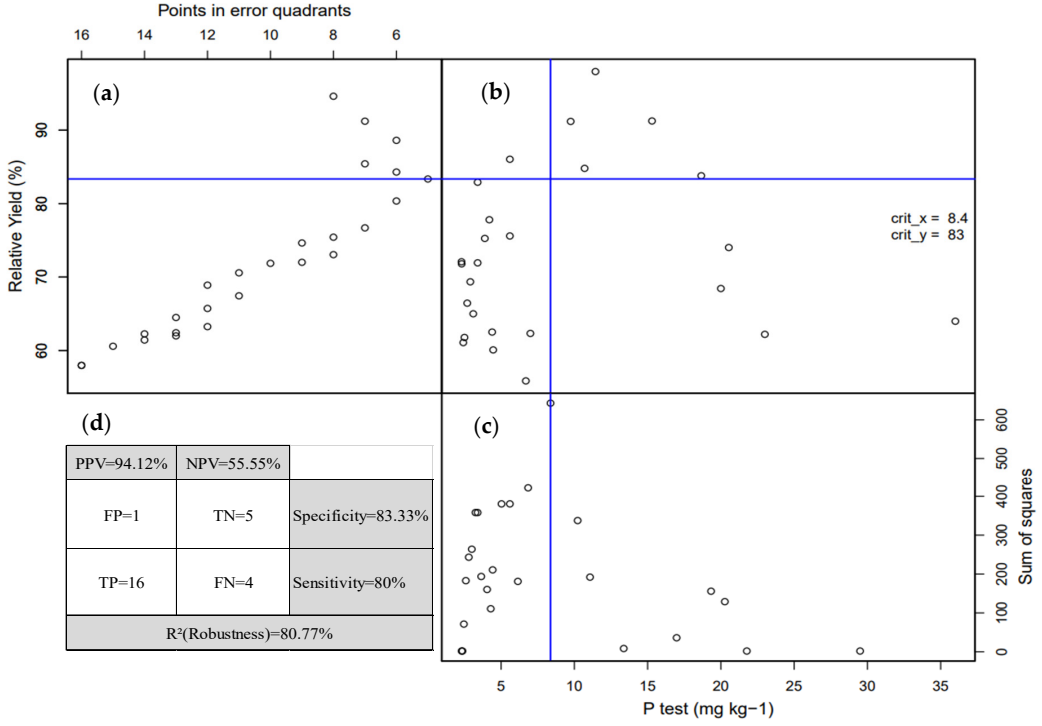

Figure 3. Cate-Nelson partitioning model for P_{Olsen} showing: (**a**) the number of points outside the model used in determining the critical relative yield (crit_y); (**b**) identified critical levels (crit_x and crit_y); (**c**) the sum of squares used in determining the critical P value (crit_x); and (**d**) model summary

table (lower left): false positive (FP), true negative (TN), false negative (FN), and true positive (TP)—number of points in the different quadrants; model performance indicators: robustness (R^2), sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV).

From these critical STP thresholds, we established seven fertility classes for the two diagnostic systems according to $P_{Bray\text{-}1}$ and P_{Olsen} (Table 3). Thus, this increasing classification of STP between P fertility classes corresponds to a decreasing probability of soybean crop response to P inputs. The likelihood of soybeans responding to P input in the very low and low fertility classes is 83–90%. In the high, very high, and extremely high fertility classes, the probability is only 0–20%. The critical STP values were taken as the lower boundaries of the medium-low fertility class (Table 3).

Table 3. Fertility classes derived from agronomic models of $P_{Bray\text{-}1}$ and P_{Olsen}.

	Fertility Classes						
	VL	**L**	**ML**	**MH**	**H**	**VH**	**E.H**
$P_{Bray\text{-}1}$ (mg kg^{-1})	0.00–3.75	3.75–7.50	7.50–11.25	11.25–15.00	15.00–18.75	18.75–30.00	>30.00
P_{Olsen} (mg kg^{-1})	0.0–4.2	4.2–8.4	8.4–12.6	12.6–16.8	16.8–21.0	21.0–33.6	>33.6

3.3. Traditional Approach to Fertilizer Recommendation Models Based on Soil Test Phosphorus

According to Figure 4, optimal P rates are much more frequent in the three fertility classes, VL, L, and ML. However, the points within each class are too scattered, so the P recommendations for soybeans vary. The classical models consider a central value per fertility class to counter this wide dispersion. Applying the medium of the optimal rates per class, we obtained for VL, L, and ML the respective recommendations of 39, 24, and 30 kg P ha^{-1} for the diagnostic system based on $P_{Brya\text{-}1}$ and 35, 35, and 26 kg P ha^{-1} for the system based on P_{Olsen}. Condensing site-by-site optimum rates into medium values resulted in a recommendation of 24–39 kg P ha^{-1} for soybean cultivation, or 2–4 times their P uptake level of 9–10 kg P ha^{-1}. The intuitive principle that the lower the STP is, the higher the P fertilizer recommendation should be, is only minimally reflected in the shape of Figure 4. The two regressions were weak, with correlation coefficients of 1.5 and 0.8% and shallow near-horizontal slopes of −0.245 and −0.213. Not surprisingly, Figure 4b shows an optimum rate of 80 kg P ha^{-1} in a high-phosphate fertility class H. Some P-rich soils respond well to phosphate fertilization if they are impaired by other factors such as compaction, runoff, or microbial inhibition. The low correlation, shallow slope, and significant scatter of points on both sides of the median in Figure 4 strongly suggest the need to account for many factors influencing soybean crop response to phosphate fertilization when making on-farm decisions. In order to achieve this, it is essential to gather and incorporate data pertaining to climatic conditions, soil properties, and management practices into a dedicated and ongoing database. By utilizing artificial intelligence, this comprehensive information can be leveraged to determine optimal P rates and effectively guide decision-making.

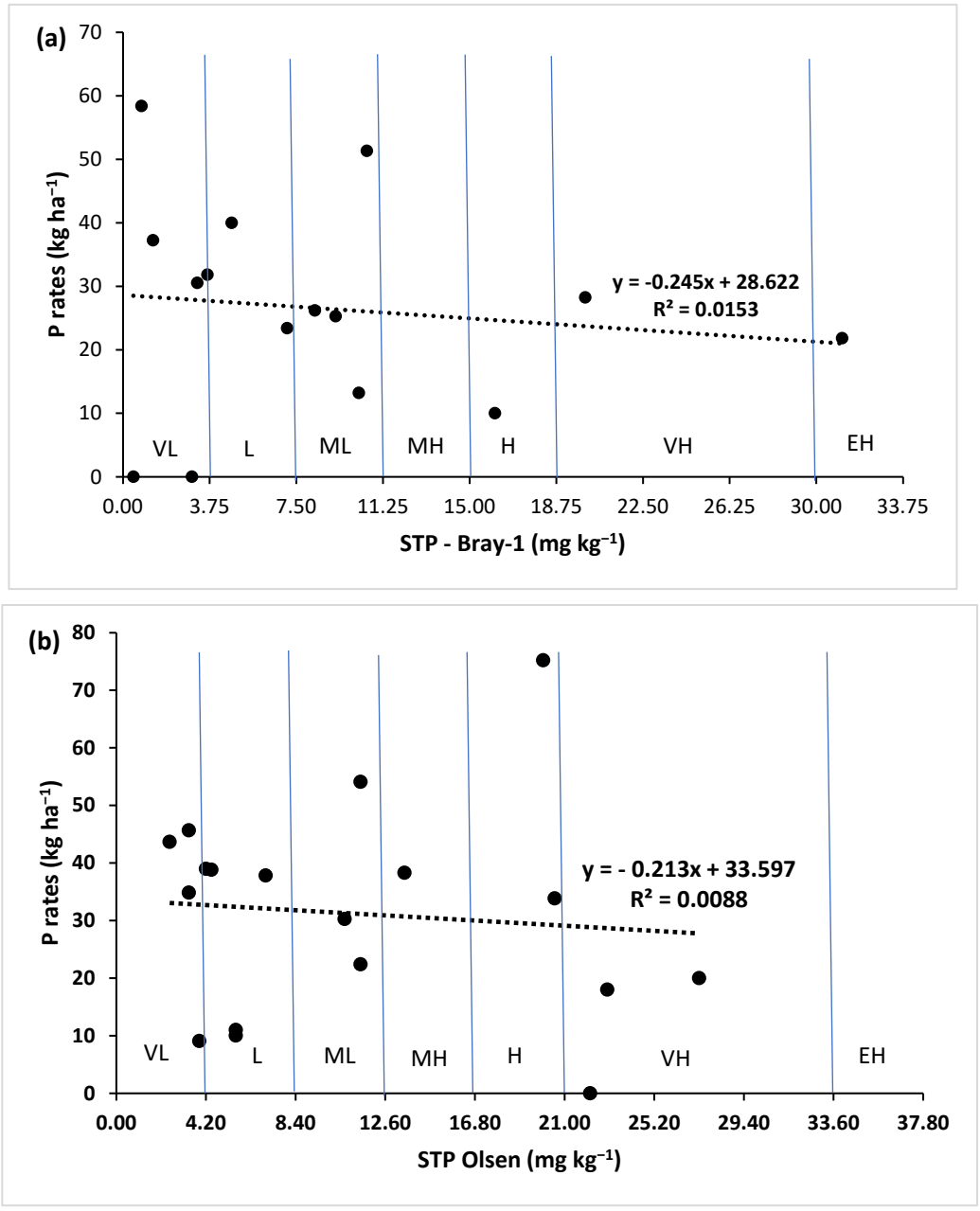

Figure 4. Fertilizer P recommendation models based on: (**a**) P_{Bray-1}, and (**b**) P_{Olsen} soil test phosphorus diagnostics encompassing seven fertility classes, including very low (VL), low (L), medium low (ML), medium high (MH), high (H), very high (VH), and extremely high (EH).

3.4. Machine Learning Prediction of Soybean Response to P Fertilizer

To predict soybean response to P fertilizer, we applied the RF model to estimate yield gain (ΔY) as a function of these seven parameters: P rates, soil pH, organic matter, STP (fertility classes), precipitation, texture, and fertilizer application methods. A linear regression between the predicted and measured ΔY gave good results about the four model performance assessment criteria. For the training set (Figure 5a), the linear regression showed a robustness (R^2) of 87.4%, RMSE of 140.81 kg ha^{-1}, slope or accuracy of 0.73, and intercept (sensitivity) of 72.05 kg ha^{-1}. These criteria for the testing set (Figure 5b) remained successful with 60.9% of R^2, 166.17 kg ha^{-1} of RMSE, 0.61 of the slope, and 96.38 kg ha^{-1} of intercept. The R^2 obtained for both training and testing sets was above the 50% target threshold of Albera et al. [28]. Similarly, the accuracies or slope obtained for both cases were closer to 1 (perfect prediction) than to 0 (null prediction). At the same time,

the RMSEs were way smaller than the mean soybean yield (i.e., 1912 kg ha^{-1}) calculated from the database used for the study and close to zero. The order of importance of variables in predicting gain yield ΔY was P rates > OM> precipitation > pH_{water} >STP > Fertilizer application method > soil texture. Therefore, we can state that aside from P rates, the three other variables that significantly influenced ΔY prediction were OM, precipitation, and soil pH_{water} (Figure 5c). The ranges for each of these four predictors are shown in Figure 5d–g. In general, the predicted yield gain increased, up to a specific limit, with the values of the predictor variables. As expected, the P rate was the most crucial variable in predicting yield gain, hence the importance of considering reconstructing response curves by RF prediction of yield gains at P rates.

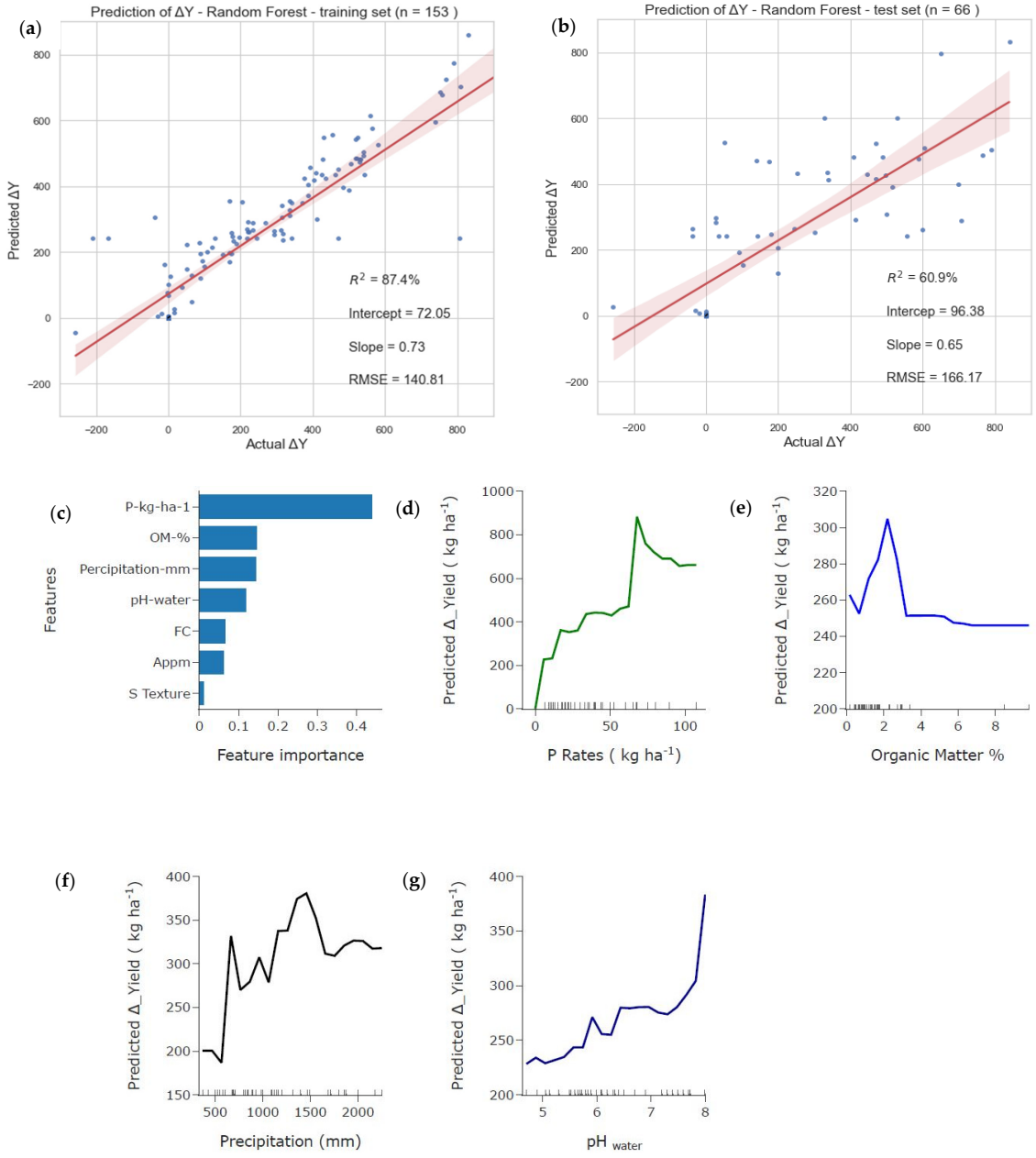

Figure 5. Gain yield (ΔY) prediction and statistical analysis for algorithm Random Forest. (**a**) Correlation between measured and predicted ΔY in the training set; (**b**) correlation between measured and predicted ΔY in the testing set; (**c**) scaled importance of predictors (P rates, organic matter, precipitation, and pH_{water}); (**d**–**g**) partial dependence on P rates, organic matter, precipitation, and pH_{water}.

3.5. Reconstruction of Soybean Response Curves to Gradual Increases in Phosphorus Rates

The RF model was employed accurately to predict the response curves at each data point. As a result, we reconstructed these response curves for experimental sites where more than two fertilizer application rates were reported. Subsequently, these curves were fitted to one of four simple mathematical models: linear, linear-plateau, quadratic, or cubic. In order to elucidate this concept further, we have included two examples of fitted response curves consisting of five data points, which were obtained from an experimental setup involving varying levels of P application (Figure 6). Figure 6a shows a typical response curve obtained from the RF for experimental site 10. Figure 6b demonstrates a highly pronounced quadratic response to phosphate fertilization. In contrast, Figure 6c exhibits no significant response. We assessed yield gain (ΔY) using RF for both figures (Figure 6b,c), assuming the same P rates. The predicted and measured curves exhibited similarities for the P-responsive site (Figure 6b), as they displayed similar shapes and fitted well to the quadratic model. When comparing the first derivatives of these models to the productivity limit (Equation (1)), the two quadratic curves indicated optimal P rates of approximately 30 and 32 kg P ha^{-1}. Conversely, for the P-unresponsive site (Figure 6c), there was no correlation between the P doses and yield gain. Consequently, the predicted and measured response patterns did not conform to a mathematical model, leading to the conclusion of zero fertilization.

Our findings indicate that approximately 50% of the soybean experimental sites exhibited a response to P fertilization, while the remaining 50% did not. Among the responsive sites, including the example depicted in Figure 6a, the curves generated through RF prediction displayed similar shapes and mathematical fitting patterns to the actual observations. The selection of mathematical models was based on the regression coefficient (R^2), which ranged from 0.8 to 1.0. In a few experimental sites, the cubic model better represented the response curves, which exhibited an S-shaped pattern and accounted for a lag effect caused by a low P rate. For the majority, the quadratic model was the most suitable. This model allows for determining fertilizer rates that maximize profitability. However, in 50% of the cases, the data points did not exhibit a clear trend (Table 4). Instead, they were scattered around a plateau, rendering no model suitable for fitting. In such instances, these sites were considered unresponsive to phosphate fertilization, and an optimal zero rate was assumed.

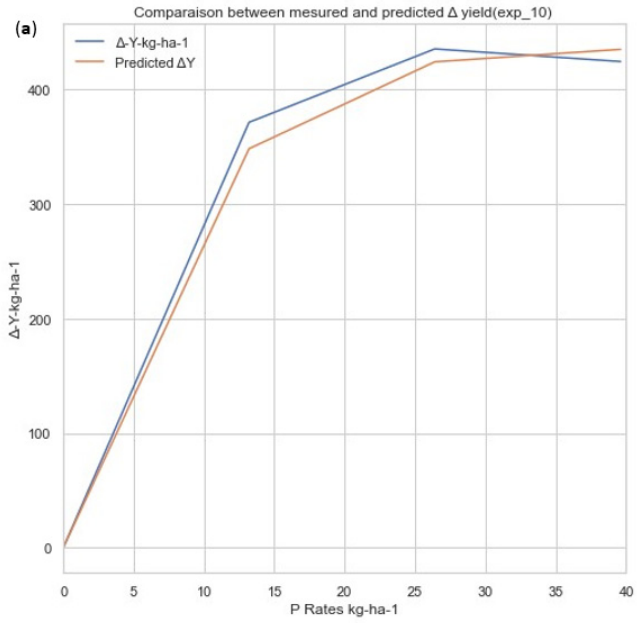

Figure 6. *Cont.*

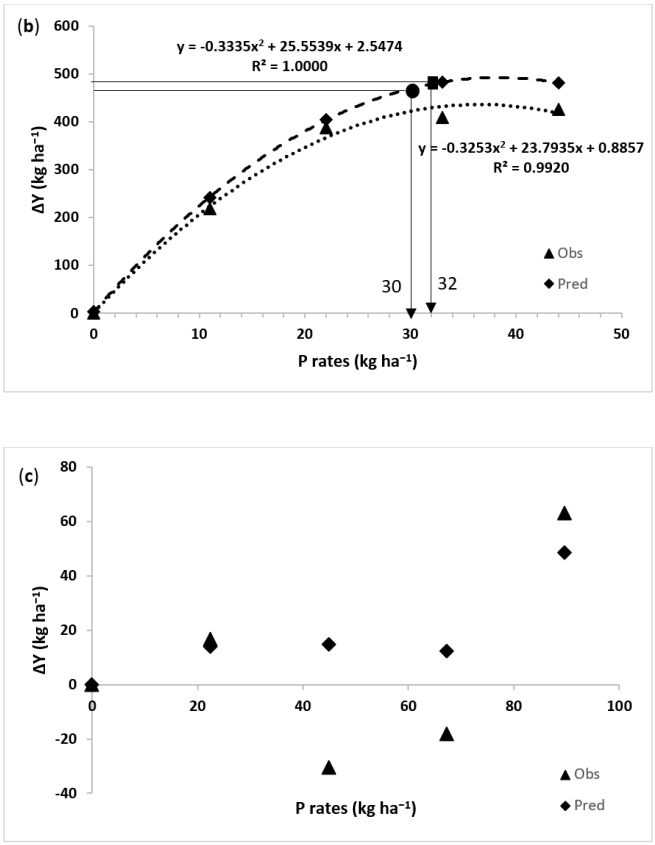

Figure 6. Example of the reconstruction the RF response curve to infer the actual and predicted optimal rates of P: (**a**) a sample response curve generated from RF, (**b**) reconstructed response curve for a P-responsive site 56, and (**c**) reconstructed curve for a P-unresponsive site 69.

Table 4. Determination of the economically optimal rate of P based on (a) the actual measured response curve, (b) the response curve predicted by the Random Forest algorithm, and (c) the traditional STP-based approach.

Optimal P Rates for P Responsive Sites with Response Curves						Optimal P Rates for P Unresponsive Sites without Response Curves					
Site	**FC**	**(a) (kg ha^{-1})**	**(b) (kg ha^{-1})**	**(c) (kg ha^{-1})**	**Model Type**	**Site**	**FC**	**(a) (kg ha^{-1})**	**(b) (kg ha^{-1})**	**(c) (kg ha^{-1})**	**Model Type**
57	VL	9	9	35	Q	16	VL	0	0	39	NR
54	VL	51	30	35	Q	60	VL	0	0	35	NR
53	VL	35	30	35	Q	61	VL	0	0	35	NR
52	VL	38	38	35	Q	62	VL	0	0	35	NR
48	VL	50	39	35	C	63	VL	0	0	35	NR
41	VL	30	30	39	C	64	VL	0	0	35	NR
27	VL	4	4	39	LP	65	VL	0	0	35	NR
24	VL	37	36	39	C	15	L	0	0	24	NR
8	VL	0	0	39	L	17	L	0	0	24	NR
6	VL	58	33	39	Q	18	L	0	0	24	NR
55	L	28	31	35	Q	19	L	0	0	24	NR
51	L	20	21	35	Q	20	L	0	0	24	NR

Table 4. *Cont.*

Optimal P Rates for P Responsive Sites with Response Curves						Optimal P Rates for P Unresponsive Sites without Response Curves					
Site	**FC**	**(a) (kg ha^{-1})**	**(b) (kg ha^{-1})**	**(c) (kg ha^{-1})**	**Model Type**	**Site**	**FC**	**(a) (kg ha^{-1})**	**(b) (kg ha^{-1})**	**(c) (kg ha^{-1})**	**Model Type**
49	L	39	33	35	Q	21	L	0	0	24	NR
45	L	38	34	35	Q	29	L	0	0	24	NR
12	L	9	9	24	LP	30	L	0	0	24	NR
7	L	40	40	24	LP	31	L	0	0	24	NR
5	L	23	20	24	C	32	L	0	0	24	NR
69	ML	0	0	26	NR	33	L	0	0	24	NR
68	ML	0	0	26	NR	34	L	0	0	24	NR
56	ML	30	32	26	Q	35	L	0	0	24	NR
26	ML	51	31	30	C	36	L	0	0	24	NR
13	ML	34	27	30	Q	37	L	0	0	24	NR
11	ML	44	25	30	Q	38	L	0	0	24	NR
10	ML	25	26	30	Q	39	L	0	0	24	NR
9	ML	21	22	30	Q	40	L	0	0	24	NR
67	H	22	22	0	LP	58	L	0	0	35	NR
47	H	43	43	0	LP	59	L	0	0	35	NR
44	H	34	36	0	Q	22	ML	0	0	30	NR
28	H	10	10	0	LP	23	ML	0	0	30	NR
3	H	54	47	0	C	14	MH	0	0	0	NR
46	VH	54	33	0	Q	66	MH	0	0	0	NR
4	VH	0	0	0	NR	43	VH	0	0	0	NR
2	VH	22	22	0	LP	42	EH	0	0	0	NR
1	VH	28	28	0	C						
50	EH	67	75	0	C						
25	EH	22	22	0	LP						

FC refers to fertility class being VL—very low, L—low, ML—medium low, MH—medium high, H—high, VH—very high, EH—extremely high. The type of model represented by Q—quadratic, C—cubic, L—linear, LP—linear-plateau, and NR—non-responsive.

Table 4 presents a comparison of optimal P fertilizer rates on a site-by-site basis using three different methods: (a) the actual response curve optimization method, which derives the measured optimal P rate; (b) the RF predicted response curve, which derives the predicted optimal P rate; and (c) the traditional STP-based P recommendation method, which provides a single optimal P rate value per fertility class. These have been presented under two categories, namely optimal P rates for P responsive sites and optimal P rates for P unresponsive sites. The optimal rates obtained from the observed and predicted response curves showed similarities (Table 4). They were equal in 71% of the cases, with a slight difference of $\leq$3 kg P ha^{-1} in 83% and differences $\geq$ 10 kg P ha^{-1} in only 9% of the cases. It is worth mentioning that in the high (H), very high (VH), and extremely high (EH) fertility classes, the traditional STP-based grid approach did not recommend any P fertilizer due to the assumption that soil test P levels were sufficient to meet the crop's P demand. In contrast, our study identified 10 response curves that could be fitted to mathematical models, leading to optimal rates ranging from 10 to 75 kg P ha^{-1} across the three fertility

classes, H, VH, and EH (Table 4). This once again highlights the limitation of the traditional STP-based fertilizer approach, which assigns a median optimal P rate for all sites within the same fertility class.

4. Discussion

4.1. Descriptive Statistics for Key Variables of Interest

Due to the wide range of soil OM levels, ranging from 0.2 to 9.8%, the response to phosphate fertilization was anticipated to vary significantly. In highly weathered soils with exceptionally low OM, soybean growers typically apply approximately twice the amount of phosphate compared to the quantity harvested in soybeans [73]. Conversely, in soils characterized by a high organic matter content, carbon mineralization, and solubilization compete with P at its fixation sites, rendering phosphorus more accessible [74]. Consequently, the necessity for mineral fertilization is reduced. The soil acidity-alkalinity conditions exhibit a high degree of variability, spanning from a very acidic pH of 4.7, which falls well below the lower limit of the optimal range (6.2–7.0) defined by CRAAQ (2010) and (5.5–7.0), as reported by Ferguson et al. [75].

Soils with pH values above 8.0 are considered distinctly alkaline, exceeding the upper limit of the optimal pH range for soybeans. These diverse pH conditions significantly influence soybean yields and their response to phosphate fertilization. The substantial range of annual rainfall, from 372 to 2249 mm, is crucial in the considerable variability observed in soybean yield and its response to P fertilizer. The influence of precipitation on soybean yields is evident. However, rainfall distribution throughout the soybean growing season can pose challenges.

Despite being situated in a region characterized by high precipitation levels, even a brief dry spell lasting a few weeks during critical phases of soybean development can lead to substantial yield losses. Regrettably, our available data on precipitation are aggregated globally, limiting our ability to analyze temporal variations on a finer scale. The average soybean yield obtained in the present study falls within the range of reported values for soybean yields in selected countries worldwide. For example, in Africa, the highest average soybean yield of 2290 kg ha^{-1} was recorded in South Africa, followed by Zambia (1940 kg ha^{-1}), Nigeria (960 kg ha^{-1}), and Uganda (600 kg ha^{-1}) [76]. Akter et al. [54] reported an average soybean yield of 1200 kg ha^{-1} for Bangladesh. Masuda and Goldsmith [77] reported average yield values of 2700 kg ha^{-1} and 1720 kg ha $^{-1}$ for the USA and China, respectively. It is worth noting that these reported values are lower than the global average yield of 2800 kg ha^{-1} [78].

4.2. Cate-Nelson Partitioning and Soil Fertility Classification

Contrary to initial expectations, applying the Cate-Nelson method yielded a critical value of $P_{Bray\text{-}1}$ of 7.5 mg kg^{-1}, lower than the critical value of P_{Olsen} of 8.4 mg kg^{-1}. The determination of these two agronomic critical values was based on the principle of maximizing the number of points in the true quadrants (TP+TN) while minimizing their occurrence in the false quadrants (FP+FN) (Figures 2 and 3). Soybean yields exhibit rapid changes in response to very low $P_{Bray\text{-}1}$ or P_{Olsen} levels within narrow ranges of 0–7.5 mg kg^{-1} and 0–8.4 mg kg^{-1}. Once the soil test levels surpass this range, soybean yields typically reach approximately 83–90% of their maximum potential. Moreover, within the STP levels falling within these ranges, the probability of response to phosphate fertilizer (PPV) was significantly higher for P_{Olsen}, at 94.12% (Figure 3), compared to $P_{Bray\text{-}1}$, which had a PPV of 65.52% (Figure 2). The calibration of $P_{Bray\text{-}1}$ is less successful than P_{Olsen}. The other three probability levels for robustness (40.35%), specificity (28.57%), and sensitivity (44.18%) are all below 50% and considerably lower than those of P_{Olsen}, which exhibited probabilities over 80% for all three measures (Figure 3). These results indicate that the $P_{Bray\text{-}1}$ method is less reliable and accurate in predicting and assessing the soybean response to phosphorus fertilization based on phosphorus availability than the P_{Olsen} method. The lower probabilities observed in the case of $P_{Bray\text{-}1}$ can be attributed to the impact of soil

properties on the P extraction capacity of this method. One potential factor is the presence of calcium carbonate ($CaCO_3$) in soil, which can neutralize all the HCl in the Bray solution. This neutralization process can lead to the formation of CaF, which inactivates the F^- and renders it ineffective in extracting P from the soil [79].

In contrast, the Olsen method, initially developed for extracting P from calcareous soils, has also been found suitable for acid soils, as Holford [80] and Sims and Ellis [81] have suggested. The critical STP values obtained in this study closely align with those reported in previous studies. Getahun et al. [3] found a critical STP value of 8.5 mg kg^{-1} of P_{Bray} in western Ethiopia. Appelhans et al. [41] also reported two critical STP values, 13.2 mg kg^{-1} and 8.5 mg kg^{-1} of $P_{Bray\text{-}1}$, for soils in the Pampa region of Argentina, sampled at depths of 0–5 cm and 0–20 cm, respectively. These consistent results in other areas indicate the general applicability of critical STP values for developing a phosphorus recommendation model for soybean crops. However, Borges and Mallarino [6] and Ferguson et al. [75] reported higher critical $P_{Bray\text{-}1}$ values of 20 mg kg^{-1} and 12 mg kg^{-1}, respectively, in Iowa and Nebraska, USA. These higher critical values indicate that phosphorus availability and response to phosphorus fertilization can vary depending on specific soil conditions, climate, and agricultural practices in different regions. It emphasizes the importance of considering these local factors when determining optimal phosphorus management strategies for soybean production. The low probabilities observed for $P_{Bray\text{-}1}$ align with the findings of earlier studies [7,8], which have indicated a considerable variation in soybean yield response to phosphorus (P) fertilization across different locations. These studies have demonstrated minimal or negligible response to P fertilization in soils with low initial P levels, while other studies have shown a substantial yield response in soils with high initial P levels [5,6].

For the critical value of P_{Olsen}, Rehm et al. [82] and Sawyer et al. [83] reported 11 mg kg^{-1} in Minnesota and 14 mg kg^{-1} in Iowa, respectively. These values are higher than the critical value of 8.4 mg kg^{-1} shown in Figure 3. These variations in critical P_{Olsen} values across studies can be attributed to several factors, including differences in soil characteristics, climate, cropping systems, and regional practices. Furthermore, Anthony et al. [84] demonstrated that alkaline soils with a pH greater than 7.2 exhibited higher critical STP values than acidic soils with a pH lower than 6.7. In addition, Watanabe and Olsen [13] recommended employing the Olsen method exclusively for alkaline soils. Nevertheless, developing a comprehensive global agronomic model that exhibits consistently high probability levels in predicting the response to phosphate fertilizer could greatly interest agronomists and farmers worldwide who lack a specific phosphate fertilizer recommendation model for soybeans. Furthermore, adopting such a model on an international scale could provide valuable guidance for optimizing phosphorus management in soybean production.

The range for the fertility classes determined for the $P_{Bray\text{-}1}$ for the current study is slightly lower than what was used by Frazen [85], who reported five fertility classes, namely 0–5 mg kg^{-1} (VL); 6–10 mg kg^{-1} (L); 11–15 mg kg^{-1} (M); 16–20 mg kg^{-1} (H) and >21 mg kg^{-1} (VH) $P_{Bray\text{-}1}$, with the VH fertility class still falling within the range determined in the present study. On the other hand, the five fertility classes and range determined by Wortmann [86] for the P_{Olsen} diagnostic included 0–3 mg kg^{-1} (VL); 4–7 mg kg^{-1} (L); 8–11 mg kg^{-1} (M); 12–15 mg kg^{-1} (H); and >15 mg kg^{-1} (VH), which also puts the VH fertility class in the range similar to what has been used for the present study.

Once the two critical values were established, we divided the soil fertility into seven classes (as shown in Table 3). First, we divided the critical value by two to delineate the low fertility classes (VL and L). Next, we multiplied the critical value by multiples of two for the medium- (ML and MH) and high- (H, VH, and EH) fertility classes. This classification scheme is based on the assumption of a nonlinear response of soybean to increasing STP levels, as proposed by Bray [87].

Our classification for $P_{Bray\text{-}1}$ is indeed slightly more specific compared to the five-class classification utilized by Frazen [85], which is as follows: 0–5 (VL), 6–10 (L), 11–15 (M),

16–20 (H), and >21 (VH) mg $P_{Bray\text{-}1}kg^{-1}$. Notably, the profitable responses mainly occur within the VL, L, and ML classes. Consequently, the agronomic interpretation based on $P_{Bray\text{-}1}$ becomes highly comparable for these first three classes. On the other hand, our second classification based on P_{Olsen} closely aligns with the five-class classification proposed by Wortmann (2018): 0–3 (VL), 4–7 (L), 8–11 (M), 12–15 (H), and >15 (VH) mg P_{Olsen} kg^{-1}. This similarity encourages us to consider internationalizing the agronomic model based on P_{Olsen} and subdividing it into seven fertility classes, as depicted in Table 3.

4.3. Traditional Approach to Fertilizer Recommendation Models Based on Soil Test Phosphorus

The establishment of fertilization grids or recommendation models, as shown in Figure 4, faces several significant challenges. These include:

i. Limited correlations exist between soil test phosphorus (STP) levels and optimal phosphorus (P) fertilizer application rates. The correlation coefficients for $P_{Bray\text{-}1}$ and P_{Olsen} are only 1.53% and 0.88%, respectively. These values indicate proximity to zero, implying a lack of substantial correlation rather than a strong positive relationship (100% correlation). Such low correlations are expected in models developed using the traditional approach. The observed weak correlations between STP levels and optimal phosphate fertilization rates align with the findings of previous studies, such as those by Mabapa et al. [36] and Morris et al. [88]. These studies have consistently demonstrated that the response of soybean crops to phosphate fertilization exhibits a high degree of unpredictability and randomness. The lack of solid correlations further emphasizes the complexity and variability in soybeans' response to phosphate fertilization. This reinforces the challenges of developing accurate fertilizer recommendation models based solely on STP levels.
ii. The representation of soil fertility levels is imbalanced, characterized by an overabundance of low-fertility soils and a scarcity of high-fertility soils. This imbalance creates challenges, particularly in high-fertility classes with limited data regarding optimal fertilizer doses. In such cases, agricultural advisors should verify the selection of fertilizer quantities while adhering to the range suggested by the model. This can be accomplished through on-farm trials and, if necessary, with the assistance of specialized services. Incorporating on-farm trials as part of soil quality monitoring allows for fine-tuning recommendations to align with local conditions. Furthermore, these trials offer an opportunity to augment the size of databases. By conducting on-farm trials, advisors refine fertilizer recommendations and contribute to the expansion of data resources for Machine Learning algorithms, ultimately enhancing the accuracy and reliability of the models used.
iii. The necessity to consider a median value arises from the wide variations observed in optimal fertilizer doses within each fertility class. A visual representation of this phenomenon can be seen in Figure 4a, specifically within the VL fertility class, where the optimal phosphorus fertilizer doses range from 0 to 58 kg P per hectare. Despite the considerable variability, a recommended median optimum value of 35 kg P ha^{-1} is determined for this specific fertility class. All the diverse optimal dose values are condensed into a single representative value using the median value. However, this approach presents challenges, resulting in under-fertilization by 23 kg P ha^{-1} at the site with an optimum dose of 58 kg. At the same time, overfertilization occurs by 35 kg at the site that did not respond to phosphorus. This example highlights the substantial variations in optimal P fertilizer doses even within a single fertility class, emphasizing the difficulties encountered when developing precise fertilization grids using the traditional approach.
iv. The occurrence of contradictory situations where high P fertilizer doses are recommended for soils with high P content, and vice versa. These inconsistencies have been highlighted in the studies done by Anthony et al. [84] and Cox et al. [7]. Considering these constraints, agronomists and knowledgeable farmers tend to be cautious when relying solely on fertilization grids developed using the traditional approach. Instead,

they prefer to incorporate their historical observations of other climatic, agronomic, and soil-related factors and make intuitive adjustments to the recommendations.

Using cumulative data from national literature has guided us in determining the phosphorus (P) doses to be applied for soybeans in different fertility classes. According to the $P_{Bray\text{-}1}$ diagnosis, the recommended P doses are 39, 24, and 30 kg P ha^{-1} for the VL, L, and ML fertility classes, respectively. Meanwhile, based on the P_{Olsen} diagnosis, the suggested P doses are 35, 35, and 26 kg P ha^{-1} for the same fertility classes. Notably, these rates align closely with the findings of Wortmann et al. [86] for the state of Nebraska, who reported P doses of 31, 19, and 10 kg P ha^{-1}, as well as the results of Grewing et al. [89] for South Dakota, who obtained P doses of 45, 28, and 8 kg P ha^{-1}. Furthermore, the maximum recommended P dose of 39 kg P ha^{-1} coincides with the findings of Awuni et al. [1].

However, despite efforts to incorporate meta-analysis and consider additional influencing factors beyond soil test phosphorus (STP), assessing P rates for soybeans remains predominantly empirical [90]. Although adaptations such as using the agri-environmental indicator PSI (Phosphorus Saturation Index) and developing new grids based on PSI have been made in agri-environmental models, further advancements in the traditional approach to developing fertilization grids have proven challenging. The second constraint of traditional grid models is that farmers receive a single recommended dose of phosphorus (P) fertilizer based on their soil's fertility class. These recommendations are often static and do not consider changes in market prices or economic considerations. As a result, farmers do not always have the opportunity to optimize their fertilization practices according to current economic conditions. Fortunately, the emergence of artificial intelligence in the past decade presents promising opportunities for a more reliable and accurate system [25]. By harnessing the power of AI, we can enhance the precision and effectiveness of fertilization strategies, paving the way for improved agricultural practices.

4.4. Prediction of P Fertilizer Rates with Machine Learning

4.4.1. The Random Forest Model and Influence of Predictor Variables

Despite significant advancements in empirical statistical and mechanistic modeling, which strive to elucidate crop response by incorporating various climatic, edaphic, and agronomic factors, accurately predicting the response of plants to phosphate fertilization remains an area that offers room for improvement in terms of error reduction [90]. Nevertheless, the remarkable aspect illustrated in Figure 5 is the unexpectedly high accuracy observed in predicting soybean response to P fertilization, which can be attributed to the utilization of artificial intelligence techniques and the meticulous selection of robust and well-calibrated models. Notably, these predictions have been achieved despite the challenges posed by data heterogeneity originating from highly variable and contrasting agro-pedoclimatic conditions.

In contrast to traditional models used to predict the appropriate dose of P to be applied (Figure 4), which yielded R^2 values ranging from 0.8 to 1.5% (essentially close to zero), the optimization through RF, considering individual sites and specific doses, they exhibited significantly improved robustness in the R^2 values, shifting away from 0% and approaching 100%. Specifically, the R^2 values for predicting the gain in soybean yield with P were 87.4% in the training dataset and 60.9% in the testing dataset. These results demonstrate that the RF model, when augmented with five other influential variables—organic matter (OM) content, annual precipitation, pH_{water}, phosphate fertility class, and the method of P fertilizer application—adequately accounts for the variations in predicting soybean yield gain (ΔY).

Furthermore, the RF model not only demonstrated improved predictive robustness but also exhibited a remarkable level of accuracy. It was observed that the model had low intercepts of 72.05 and 96.38 kg of yield gain per hectare for the training and testing datasets, respectively, which accounted for only 3.8–5.0% of the average soybean yield of 1912 kg ha^{-1} (Table 2). This minimal deviation for P doses approaching zero yield gain indicates the model's inherent flexibility. As measured by the RMSE, the overall error

was 140.81 and 166.17 kg of yield gain per hectare for the training and testing datasets, respectively, representing only 7.4–8.7% of the average soybean yield. This error level in predicting yield gain (ΔY) at P rates remains below 10%, which is highly satisfactory compared to traditional models. The yield gain prediction performances observed in this study align with the findings reported by Khaki and Wang [91]. Their study utilized meteorological variables and soil information as inputs for a Deep Neural Networks model, which yielded promising results in predicting crop yields. In addition to this study, other recent research studies have further substantiated the predictive capabilities of Machine Learning in crop yield prediction. Several past studies have demonstrated the effectiveness of incorporating climate, soil characteristics, and agronomic practices parameters in artificial intelligence models for accurate crop yield predictions [3,25,30]. These studies collectively contribute to the growing body of evidence supporting the utilization of artificial intelligence in improving crop yield forecasting.

In addition to its predictive capabilities, RF also provides valuable metrics, such as Feature Importance (Figure 5c), which greatly assisted us in interpreting the model. These metrics helped us identify the most influential factors related to soil, climate, and fertilization practices that predict the response of soybean to phosphorus addition. According to the ranking presented in Figure 5c, the order of influence is as follows: P rates > OM > precipitation > pH_{water} > fertility class > P application method > Texture. Nevertheless, significant attention should be directed toward the first four factors as they emerge as the most influential determinants. The yield gain (ΔY), which compares the yield of the P-fertilized unit with that of the non-fertilized unit, logically identifies the P rates as the primary factor in determining this response (Figure 5d). This finding aligns with the results obtained by Abera et al. [29], who observed a similar trend where P rate was the most crucial yield-influencing factor for wheat. The substantial increase in ΔY is predominantly observed within the range from 10 to 40 kg P ha-1 (Figure 5d). Beyond this range, data become less prevalent, necessitating cautious interpretations due to the potential for excessive phosphorus accumulation in the soil, which can lead to saturation and an increased risk of eutrophication in nearby surface waters [19,28]. A site-specific predictive model, as Coulibali et al. [28] suggested, would offer a viable solution to address this environmental concern.

The significance of soil organic matter cannot be overstated, as it plays a pivotal role in determining the response of soybean to phosphorus (P) addition. This observation underscores that soils with low organic matter content of 0.2–2.5% (Figure 5e) exhibit reduced biological activity, necessitating higher phosphorus levels. This aligns with conventional wisdom, as inactive soils demand substantially greater phosphorus inputs than their active counterparts to achieve equivalent yield. In addition, the significance of annual rainfall suggests that the amount of precipitation received also plays a crucial role in predicting the response of soybeans to phosphorus (P) addition. Fields that receive more than 700 mm of rainfall tend to provide optimal nourishment to the crop and can accommodate higher levels of phosphorus, resulting in substantial increases in soybean yield (Figure 5f). This observed trend of ΔY (yield gain) increases with higher rainfall levels above 700 mm is consistent with findings from previous studies, which have demonstrated that water stress or drought can impede soil phosphorus diffusion and plant phosphorus uptake, ultimately leading to reduced yields [92–94].

Regarding the fourth influencing factor, pH (Figure 5g), its range of influence on ΔY falls between 5.5 and 7.0. Below this range, the response of soybeans to P addition is weak. When the pH drops below 5.5, the Al^{3+} form becomes apparent, leading to aluminum toxicity and hindering root activity in crop plants [90]. Several studies, including those by Adams et al. [5], and Ferguson et al. [75], have confirmed that soybeans exhibit optimal performance in soils with pH values within this range. The order of importance of the last two predictors aligns with the observation made by Gupta et al. [95] that precipitation holds greater significance than soil pH in influencing crop yield within the pH range of 5.5–7.0.

This study's RF (Random Forest) model, supported by extensive data collection and comprehensive documentation of experimental P dosage sites, exhibits exceptional capabilities in generating highly personalized dose-by-dose response curves for individual diagnosed soybean fields. Furthermore, it demonstrates the ability to evaluate the stability of these curves under various scenarios, including climate change and alterations in soil properties such as organic matter depletion or acidification. These findings highlight the importance of AI models in precision fertilization, and recent studies by Coulibali et al. [28] and Lima-Neto et al. [96] further emphasize the informative potential of AI in assessing yield and dosage. These advancements contribute to the growing body of research that underscores the valuable role of AI in optimizing fertilization practices and enhancing agricultural productivity.

4.4.2. Reconstruction of Soybean Response Curves to Gradual Increases in Phosphorus Rates

A P addition response curve, tailored to newly diagnosed edaphic conditions and prevailing climatic conditions, can be used by leveraging the soybean database and applying a dose-by-dose RF predictive model. By considering the yield gains resulting from P additions and incorporating up-to-date unit prices for fertilizers and soybean production, it becomes possible to determine updated optimal P doses specific to each field. To evaluate the effectiveness of this novel approach in reconstructing response curves to P addition, it is crucial to compare the predicted response curves with the actual response curves. Furthermore, studying the similarity between the actual and predicted optimal doses validates the model's performance. This analysis allows researchers and practitioners to assess the accuracy and reliability of the RF model in capturing the intricate relationship between P addition and soybean yield gain, thus facilitating the identification of optimal P dosage recommendations for specific field conditions.

In most cases, the P doses derived from actual and RF-predicted response curves are found to be similar to, but generally lower than, those obtained using the traditional single median dose per fertility class method. This observation suggests that the actual and predicted response curves provide more accurate and tailored P dosage recommendations than the traditional approach. Table 4 presents evidence that the traditional approach tends to overestimate the doses of P to be applied in many cases. This discrepancy is particularly notable when the response curves exhibit diffuse patterns that cannot be accurately fitted to any mathematical model, indicating a lack of responsiveness to P addition. In such cases, the traditional approach may recommend excessive P doses ranging from 24 to 39 kg P ha^{-1} in 50% of instances.

These findings align with previous studies [7,85], which have also demonstrated the limitations of the traditional approach in accurately prescribing optimal fertilizer doses. The inability of the traditional method to account for the variability in field-specific conditions and accurately predict the response to P addition can lead to unnecessary over-application of P fertilizers. The results presented in Table 4 emphasize the significance of embracing advanced and individualized approaches, such as artificial intelligence (AI) prediction, in order to optimize P dosage recommendations. Using RF algorithms based on a reliable and comprehensive database can reduce uncertainty regarding optimal fertilizer dosage and minimize speculation regarding safety overdoses [97]. By incorporating multiple factors and leveraging Machine Learning techniques, it becomes possible to reconstruct response curves and fit them into simple mathematical models that are easily understandable by agronomists and farmers. This approach enables the derivation of more accurate and tailored recommendations for optimizing P fertilization, as illustrated in Figure 6a. Overall, integrating AI prediction, comprehensive databases, response curve reconstruction, and mathematical modeling facilitates the delivery of precise and field-specific recommendations for soybean P fertilization. By leveraging these advanced approaches, we can optimize crop productivity while minimizing potential excesses and environmental impacts, promoting sustainable fertilization practices.

The second significant advantage of using the RF-based approach to reconstruct response curves to P addition lies in its ability to incorporate real-time market prices for fertilizers and products. This feature allows for the deduction of the accurate optimum P dose that can be recommended to farmers to ensure profitability in soybean cultivation. Considering current market values, the approach enables more informed decision-making regarding larger investments in phosphate fertilizers for P-responsive fields. Recommendations for higher P doses would only be advised when the anticipated market value of soybean grains is attractive enough to yield a reasonable return on investment.

In contrast, the traditional approach based on a simple grid with soil test P (STP) intervals and their corresponding recommended doses lacks the ability to conduct this microeconomic analysis for farmers, considering updated phosphate input and soybean production prices. While the microeconomic analysis is carried out on a site-by-site basis in the traditional approach, the amalgamation of all optimum doses into a single recommended median dose renders the use of response curves unnecessary and readjustment of the median optimum P amounts impractical. Furthermore, the results obtained through the traditional approach are often outdated and not readily available to grid users, making it challenging to readjust the optimum P rates following the current prices of P inputs and the soybean market. We can overcome these limitations by utilizing the RF-based approach, considering real-time market prices, and providing more accurate and up-to-date recommendations for optimal P rates. Integrating economic factors into the decision-making process empowers farmers to make informed choices that align with their financial interests, promoting sustainable fertilization practices while maximizing profitability in soybean production.

5. Conclusions

In this study, we conducted a comprehensive data extraction exercise from reliable journals in the international literature. We focused on peer-reviewed studies investigating soybean response to phosphate fertilization in various agro-pedo-climatic conditions worldwide. This data extraction process resulted in a database of 219 records, encompassing 69 distinct experiments. The primary objective of our study was to compare two distinct approaches: the traditional STP-based approach and the novel response curve prediction approach, to develop an effective phosphorus recommendation system for soybean cultivation. The traditional STP-based approach involved conventional steps such as correlation, calibration, optimization, and discretization into medians based on fertility classes. In contrast, the response curve approach employed artificial intelligence techniques, specifically the Random Forest (RF) algorithm, to predict response curves to P addition. Now, let us delve into the main findings obtained from the two approaches:

- In the traditional STP-based approach, the calibration step involved using the Cate-Nelson partitioning procedure with two soil test P (STP) diagnostic systems, specifically $P_{Bray\text{-}1}$ and P_{Olsen}. This calibration process identified critical values of 7.5 and 8.4 mg kg^{-1} for $P_{Bray\text{-}1}$ and P_{Olsen}, respectively. These critical values serve as thresholds above which the response to phosphate fertilizers becomes relatively weak. Based on these critical values, the seven phosphate fertility classes for soybeans were delineated, ranging from very low (VL) to extremely high (EH). Subsequently, a site-by-site optimization was performed, followed by discretization into median values of the optimal doses within each of the seven fertility classes. This process resulted in the development of two traditional P recommendation grids. The recommended doses for the $P_{Bray\text{-}1}$-based diagnostic system for the first three fertility classes (VL, L, and ML) were 39, 24, and 30 kg P ha^{-1}, respectively. In the case of the P_{Olsen}-based system, the corresponding recommended doses for the same three classes were 35, 35, and 26 kg P ha^{-1}. In the higher fertility classes (MH, H, VH, and EH), the recommendations were limited to a fixed P uptake value of 10 kg P ha^{-1}. These two classic grids align closely with those found in various fertilization guides commonly available. However, it is worth noting that the correlation between soil test phosphorus (STP)

levels and the corresponding optimal phosphorus (P) doses in these grids was relatively low. The correlation coefficient was only 1.53% for $P_{Bray\text{-}1}$ and 0.88% for P_{Olsen}. Despite the limited correlation coefficients, it is important to remember that traditional STP-based grids have been widely used and accepted in practical agricultural settings. These grids provide general guidelines for phosphorus fertilization based on soil fertility classes. However, the low correlation coefficients indicate that the STP levels alone may not accurately predict optimal P doses.

- In both the traditional and artificial intelligence approaches, the optimization step included fitting the response curves to P addition using simple mathematical models. This process aimed to determine the economically optimal dose of P for each soybean site. The optimization results revealed that approximately 50% of the experimental soybean sites responded positively to P fertilization, indicating that adding P significantly impacted yield. On the other hand, the remaining 50% of the sites did not exhibit a significant response to P fertilization. This finding highlights the importance of considering site-specific factors and individualized approaches when recommending P fertilization.
- The new artificial intelligence approach using Random Forest (RF)-based optimization has demonstrated high robustness in both the training and testing datasets. With 87.4% robustness in the training dataset and 60.9% in the testing dataset, the RF model accurately predicted the yield gain (ΔY) from adding phosphorus (P), with prediction error levels below 10%. This level of accuracy is very satisfactory when compared to traditional models. By incorporating five influential variables (organic matter content, annual rainfall, pH-water, phosphate fertility class, and P fertilizer application method), the RF model successfully generated response curves that exhibited similar shapes and fitting patterns to the actual curves. This indicates that the RF model was able to capture the complex relationships between these variables and the response to P addition, leading to comparable dose recommendations. Furthermore, the comparison between the optimal P doses obtained from the observed response curves and those predicted by the RF model showed a high level of agreement. In 71% of cases, the observed and predicted optimum doses were equal, and in 83% of cases, the difference was $\leq$3 kg P ha^{-1}. Larger differences (>10 kg P ha^{-1}) were observed in only 9% of cases. This demonstrates the consistency and accuracy of the RF model in reconstructing response curves and providing specific prescriptions for optimal P rates in soybean fields. Overall, these results highlight the effectiveness and reliability of the artificial intelligence optimization approach using RF-based modeling. The approach surpasses traditional models in terms of accuracy, specificity, and precision in predicting optimal P doses to be prescribed for soybean cultivation. One of the notable advantages of RF-predicted response curves is their ability to incorporate real-time market prices for fertilizers and crop products. By integrating economic factors into the decision-making process, farmers can make informed choices that align with their financial interests, promoting sustainable fertilization practices tailored to their specific characteristics of soybean fields. Farmers can determine the economic feasibility of applying different phosphorus (P) fertilizer doses by considering current market prices. This information allows them to make decisions that optimize crop productivity and profitability. If the market value of soybean grains is high, farmers may be more inclined to invest in higher P doses, as the expected returns on investment are more attractive. On the other hand, if market conditions are less favorable, farmers can adjust their P dosage accordingly to minimize unnecessary expenses. Indeed, traditional single fertilization grid models do not offer the flexibility to incorporate real-time economic factors for readjustment. These models typically provide fixed fertilizer recommendations based on predetermined fertility classes and do not consider the dynamic nature of market prices for fertilizers and crop products.
- The use of several variables aside from STP in the RF to predict soybean yields and deduce optimal P fertilizer application rates validates that it is a more accurate

approach which does not simplify reality since crop response to fertilization depends on many other factors (edaphic, climatic, and management practices).

Although the present study has shown that the Random Forest can be used to accurately predict soybean yields to deduce optimal P fertilizer application rates, the novelty approach of Machine Learning (ML) and its applicability in agriculture has some limitations. One limitation is the difficulty in coping with model uncertainty when using complicated data-model fusion algorithms for yield prediction. Another limitation is that developing an accurate and effective model for crop yield predictions require considering various factors such as edaphic, crop diseases, management practices, climate data, and crop classification based on development phase, which is always challenging to obtain and integrate. The accuracy of the predictions is also strongly dependent on the size of the database, i.e., the larger the database, the better the results. Furthermore, the accuracy of ML techniques for estimating crop yields can vary depending on the algorithm used. While Random Forest has shown high accuracy in some studies, other studies have used other algorithms and obtained fairly good results. Therefore, the choice of ML algorithm for yield prediction should be carefully considered based on the specific requirements and limitations of the study.

Author Contributions: Conceptualization, L.K. and F.M.C.; methodology, L.K. and F.M.C.; writing—original draft, F.M.C.; Machine Learning, H.J., L.K. and I.K.; project administration, M.I., writing—review and editing, L.K. and M.I.; supervision, L.K. and M.I.; funding acquisition, L.K. All authors have read and agreed to the published version of the manuscript.

Funding: OCP-Group provided funding through the P-Management on a Global Scale Project, AS No. 120.

Data Availability Statement: The raw data used to assemble this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that this study received funding from OCP-Group. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Awuni, G.A.; Reynolds, D.B.; Goldsmith, P.D.; Tamimie, C.A.; Denwar, N.N. Agronomic and economic assessment of input bundle of soybean in moderately acidic Savanna soils of Ghana. *Agrosystems Geosci. Environ.* **2020**, *3*, e20085. [CrossRef]
2. Buczko, U.; Van Laak, M.; Eichler-Löbermann, B.; Gans, W.; Merbach, I.; Panten, K.; Peiter, E.; Reitz, T.; Spiegel, H.; Von Tucher, S. Re-evaluation of the yield response to phosphorus fertilization based on meta-analyses of long-term field experiments. *Ambio* **2017**, *47*, 50–61. [CrossRef]
3. Getahun, D.E.; Dereje, A.L.; Bekel, A.N.; Tigist, A.D. Soil test based Phosphorous Calibration for Soybean [*Glycine max* (L.) Merrill] Production on Nitisols in Assosa Zone of Benishangul Gumuz Region, Western Ethiopia. *Greener J. Soil Sci. Plant Nutr.* **2018**, *5*, 23–30.
4. Heard, J.; Hay, D. Typical nutrient content, uptake pattern and carbon: Nitrogen ratios of prairie crops. Designing Cropping Systems that Prosper in Variable Weather. In Proceedings of the 7th Manitoba Agronomists Conference, Winnipeg, MB, Canada, 12–13 December 2006.
5. Adams, J.F.; Adams, F.; Odom, J.W. Interaction of Phosphorus Rates and Soil pH on Soybean Yield and Soil Solution Composition of Two Phosphorus-Sufficient Ultisols. *Soil Sci. Soc. Am. J.* **1982**, *46*, 323–328. [CrossRef]
6. Borges, R.; Mallarino, A.P. Grain yield, early growth, and nutrient uptake of no-till soybean as affected by phosphorus and potassium placement. *Agron. J.* **2000**, *92*, 380–388.
7. Cox, M.S.; Gerard, P.D.; Wardlaw, M.C.; Abshire, M.J. Variability of selected soil properties and their relationship with soybean yield. *Agron. J.* **2003**, *67*, 1296–1302. [CrossRef]
8. Holzapfel, C.; Hnatowich, G.; Pratchler, J.; Webber, J.; Flaten, D. *Developing Phosphorus Management Recommendations for Soybeans in Saskatchewan*; Saskatchewan Pulse Crop Development Board Saskatoon: Saskatoon, SK, Canada, 2017; 24p.
9. Dahnke, W.C.; Olson, R.A. Soil test correlation calibration recommendation. In *Soil Testing and Plant Analysis*, 3rd ed.; Westerman, R.L., Ed.; SSSA Book Series 3; SSSA: Madison, WI, USA, 1990; pp. 45–71.
10. Wuenscher, R.; Unterfrauner, H.; Peticzka, R.; Zehetner, F. A comparison of 14 soil phosphorus extraction methods applied to 50 agricultural soils from Central Europe. *Plant Soil Environ.* **2015**, *61*, 86–96. [CrossRef]
11. Bray, R.H.; Kurtz, L.T. Determination of total. organic, and available forms of phosphorus in soils. *Soil Sci.* **1945**, *59*, 39–45. [CrossRef]

12. Mehlich, A. Mehlich 3 soil test extractant: A modification of the Mehlich 2 extractant. *Commun. Soil Sci. Plant Anal.* **1984**, *15*, 1409–1416. [CrossRef]
13. Watanabe, F.S.; Olsen, S.R. Test of an Ascorbic Acid Method for Determining Phosphorus in Water and $NaHCO_3$ Extracts from the Soil. *Soil Sci. Soc. Am. J.* **1965**, *29*, 677–678. [CrossRef]
14. Mallarino, A.P.; Blackmer, A.M. Comparison of methods for determining critical concentrations of soil test phosphorus for corn. *Agron. J.* **1992**, *84*, 850–856. [CrossRef]
15. Cope, J.T., Jr.; Rouse, R.D. Interpretation of Soil Test Results. In *Soil Testing and Plant Analysis*; Revised Edition; Walsh, L.M., Beaton, J.D., Eds.; Soil Science Society of America: Madison, WI, USA, 1973; pp. 25–54.
16. Cate, R.B.; Nelson, L.A. A simple statistical procedure for partitioning soil test correlation data into two classes. *Soil Sci. Soc. Am. Proc.* **1971**, *35*, 658–659. [CrossRef]
17. Nelson, L.A.; Anderson, R.L. Partitioning of soil testcrop response probability. In *Soil Testing: Correlating and Interpreting the Analytical Results*; ASA Special Publication 29; Stelly, M., Ed.; ASA: Madison, WI, USA, 1984; pp. 19–38.
18. Rosa, A.T.; Ruiz Diaz, D.A.; Hansel, F.D. Phosphorus fertilizer optimization is affected by soybean varieties and placement strategy. *J. Plant Nutr.* **2020**, *43*, 2336–2349. [CrossRef]
19. Khiari, L.; Parent, L.E.; Pellerin, A.; Alimi, A.R.A.; Tremblay, C.; Simard, R.R.; Fortin, J. An agri-environmental phosphorus saturation index for acid coarse-textured soils. *J. Environ. Qual.* **2000**, *29*, 1561–1567, Erratum in *J. Environ. Qual.* **2000**, *29*, 2052. [CrossRef]
20. Pellerin, A.; Parent, L.E.; Tremblay, C.; Fortin, J.; Tremblay, G.; Landry, C.P.; Khiari, L. Agri-environmental models using Mehlich-III soil phosphorus saturation index for corn in Quebec. *Can. J. Soil Sci.* **2006**, *86*, 897–910. [CrossRef]
21. Cox, F.R. Economic phosphorus fertilization using a linear response and plateau function. *Commun. Soil Sci. Plant Anal.* **1996**, *27*, 531–543. [CrossRef]
22. Waugh, D.L.; Cate, R.B., Jr.; Nelson, L.A.; Manzano, A. New concepts in biological and economical interpretation of fertilizer response. In *Seminar on Soil Management and the Development Process in Tropical America*; University Consortium on Soils of the Tropics: Raleigh, NC, USA, 1974.
23. Bardella, G. Phosphorus Management Practices for Soybean Production in Manitoba. Master's Thesis, University of Manitoba, Winnipeg, MB, USA, 2016; 197p.
24. Boring, T.J.; Thelen, K.D.; Board, J.E.; De Bruin, J.L.; Lee, C.D.; Naeve, S.L.; Ross, W.J.; Kent, W.A.; Ries, L.L. Phosphorus and potassium fertilizer application strategies in corn-soybean rotations. *Agronomy* **2018**, *8*, 195. [CrossRef]
25. Dupré, R.L.C.; Khiari, L.; Gallichand, J.; Joseph, C.A. Multi-factor diagnostic and recommendation system for boron in neutral and acidic soils. *Agronomy* **2019**, *9*, 410. [CrossRef]
26. Burke, W.J.; Frossard, E.; Kabwe, S.; Jayne, T.S. Understanding Fertilizer Adoption and Effectiveness on Maize in Zambia. *Food Policy* **2019**, *86*, 101721. [CrossRef]
27. Bondre, D.A.; Mahagaonkar, S. Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms. *Int. J. Eng. Appl. Sci. Technol.* **2019**, *4*, 371–376. [CrossRef]
28. Coulibali, Z.; Cambouris, A.N.; Parent, S.É. Site-specific machine learning predictive fertilization models for potato crops in Eastern Canada. *PLoS ONE* **2020**, *15*, e0230888. [CrossRef] [PubMed]
29. Abera, W.; Tamene, L.; Tesfaye, K.; Jiménez, D.; Dorado, H.; Erkossa, T.; Kihara, J.; Ahmed, J.S.; Amede, T.; Ramirez-Villegas, J. A data-mining approach for developing site-specific fertilizer response functions across the wheat-growing environments in Ethiopia. *Exp. Agric.* **2022**, *58*, e9. [CrossRef]
30. Manoj, K.D.; Malyadri, N.; Srikanth, M.S.; Ananda, J.B. A Machine Learning model for Crop and Fertilizer Recommendation. *NVEO-Nat. Volatiles Essent. Oils J.* **2021**, *8*, 10531–10539.
31. Farmaha, S.B.; Fernández, G.F.; Nafziger, D.E. No-till and strip-till soybean Production with surface and subsurface phosphorus and potassium fertilization. *Agron. J.* **2011**, *103*, 1862–1869. [CrossRef]
32. Appiah, A.K.; Helget, R.; Xu, Y.; Wu, J. Response of soybean yield and yield components to phosphorus fertilization in south Dakota. *Conf. Appl. Stat. Agric.* **2014**, *2014*. [CrossRef]
33. Barbagelata, B.P.; Melchiori, R.; Paparotti, O. Phosphorus Fertilization of Soybeans in Clay Soils of Entre Ríos Province. *Better Crops Int.* **2002**, *16*, 3–5.
34. Fageria, N.K.; Moreira, A.; Castro, C. Response of soybean to phosphorus fertilization in Brazilian Oxisol. *Commun. Soil Sci. Plant Anal.* **2011**, *42*, 2716–2723. [CrossRef]
35. Olaniyan, A.; Udo, E.; Afolami, A. Performance of Soybean (*Glycine Max* L.) Influenced by Different Rates and Sources of Phosphorus Fertilizer in South-West Nigeria. *Agrofor* **2016**, *1*. [CrossRef]
36. Mabapa, P.M.; Ogola, J.B.O.; Odhiambo, J.J.O.; Whitbread, A.; Hargreaves, J. Effect of phosphorus fertilizer rates on growth and yield of three soybean (*Glycine max*) cultivars in Limpopo Province. *Afr. J. Agric. Res.* **2010**, *5*, 2653–2660.
37. Pal, U.R.; Olufajo Nnadi, L.A. Response of soybean (*Glycine max* (L.) Merrill) to phosphorus, Potassium and Mdybdenum applications. *J. Agric. Sci.* **1989**, *112*, 131–136. [CrossRef]
38. Jain, P.C.; Trivedi, S.K. Response of soybean [*Glycine max* (L.) Merrill] to phosphorus and biofertiuzers. *Legume Res.* **2005**, *28*, 30–33.
39. Singh, A.K.; Singh, S. Response of soybean to phosphorus and boron fertilization in acidic upland soil of Nagaland. *J. Indian Soc. Soil Sci.* **2012**, *60*, 167–170.

40. Dabesa, A.; Tana, T. Response of Soybean (*Glycine Max* L. (Merrill)) to Bradyrhizobium Inoculation, Lime, and Phosphorus Applications at Bako, Western Ethiopia. *Int. J. Agron.* **2021**, *2021*, 6686957. [CrossRef]
41. Appelhans, S.C.; Melchiori, R.J.; Barbagelata, P.A.; Novelli, L.E. Assessing organic phosphorus contributions for predicting soybean response to fertilization. *Soil Sci. Soc. Am. J.* **2016**, *80*, 1688–1697. [CrossRef]
42. Akpalu, M. Phosphorus Application Rhizobia Inoculation on Growth Yield of Soybean (*Glycine max*, L. Merrill). *Am. J. Exp. Agric.* **2014**, *6*, 674–685. [CrossRef]
43. Ulzen, J.; Abaidoo, R.C.; Ewusi-Mensah, N.; Masso, C. On-farm evaluation and determination of sources of variability of soybean response to Bradyrhizobium inoculation and phosphorus fertilizer in northern Ghana. *Agric. Ecosyst. Environ.* **2018**, *267*, 23–32. [CrossRef] [PubMed]
44. Bhangoo, M.S.; Albritton, D.J. Effect of Fertilizer Nitrogen, Phosphorus, and Potassium on Yield and Nutrient Content of Lee Soybeans. *Agron. J.* **1972**, *64*, 743–746. [CrossRef]
45. Adjei-Nsiah, S.; Alabi, B.U.; Ahiakpa, J.K.; Kanampiu, F. Response of grain legumes to phosphorus application in the Guinea savanna agro-ecological zones of Ghana. *Agron. J.* **2018**, *110*, 1089–1096. [CrossRef]
46. Adjei-Nsiah, S.; Martei, D.; Yakubu, A.; Ulzen, J. Soybean (*Glycine max*, L. Merrill) responds to phosphorus application and rhizobium inoculation on Acrisols of the semi-deciduous forest agro-ecological zone of Ghana. *PeerJ* **2022**, *10*, e12671. [CrossRef]
47. Chiezey, U.F.; Odunze, C. Soybean response to application of poultry manure and phosphorus fertilizer in the sub-humid savanna of Nigeria. *J. Ecol. Nat. Environ.* **2009**, *1*, 25–31.
48. Houngnandan, H.B.; Adandonon, A.; Akplo, T.M.; Zoundji, C.C.; Kouelo, A.F.; Zeze, A.; Houngnandan, P.; Bodjrenou, R.; Dehoue, H.; Akinocho, J. Effect of rhizobial inoculation combined with phosphorus fertilizer on nitrogen accumulation, growth and yield of soybean in Benin. *J. Soil Sci. Environ. Manag.* **2020**, *11*, 153–163. [CrossRef]
49. Jahangir, A.; Mondal, R.; Nada, K.; Sarker, M.; Moniruzzaman, M.; Hossain, M. Response of Different Level of Nitrogen and Phosphorus on Grain Yield, Oil Quality and Nutrient Uptake of Soybean. *Bangladesh J. Sci. Ind. Res.* **1970**, *44*, 187–192. [CrossRef]
50. Pascal, U.N. Yield Parameters and Stability of Soybean [*Glycine Max*. (L.) Merril] as Influenced by Phosphorus Fertilizer Rates in Two Ultisols. *J. Plant Breed. Crop Sci.* **2015**, *5*, 54–63. [CrossRef]
51. Mahamood, J.; Abayomi, Y.A.; Aduloju, M.O. Comparative growth and grain yield responses of soybean genotypes to phosphorous fertilizer application. *Afr. J. Biotechnol.* **2009**, *8*, 1030–1036.
52. Devi, K.N.; Singh, L.N.; Devi, T.S.; Devi, H.N.; Singh, T.B.; Singh, K.K.; Singh, W.M. Response of Soybean [*Glycine max* (L.) Merrill] to Sources and Levels of Phosphorus. *J. Agric. Sci.* **2012**, *4*, 44. [CrossRef]
53. Zoundji, C.C.; Houngnandan, P.; Amidou, M.H.; Kouelo, F.A.; Toukourou, F. Inoculation and phosphorus application effects on soybean [*Glycine max* (L.) Merrill] productivity grown in farmers' fields of Benin. *J. Anim. Plant Sci.* **2015**, *25*, 1384–1392.
54. Akter, F.; Islam, N.; Shamsuddoha, A.T.M.; Bhuiyan, M.S.I.; Shilpi, S. Effect of Phosphorus and Sulphur on Growth and Yield of Soybean (*Glycine Max* L.). *Int. J. Bio-Resour. Stress. Manag.* **2013**, *4*, 556–561.
55. Shahid, M.Q.; Saleem, M.F.; Khan, H.Z.; Anjum, S. Performance of soybean (*Glycine max* L.) under different phosphorus levels and inoculation. *Pak. J. Agric. Sci.* **2009**, *46*, 237–241.
56. Begum, M.A.; Islam, M.A.; Ahmed, Q.M.; Islam, M.A.; Rahman, M.M. Effect of nitrogen and phosphorus on the growth and yield performance of soybean. *Res. Agric. Livest. Fish.* **2015**, *2*, 35–42. [CrossRef]
57. Khanam, M.; Islam, M.; Ali, M.; Chowdhury, I.F.; Masum, S. Performance of Soybean Under Different Levels of Phosphorus and Potassium. *Bangladesh Agron. J.* **2016**, *19*, 99–108. [CrossRef]
58. Abbasi, M.K.; Tahir, M.M.; Azam, W.; Abbas, Z.; Rahim, N. Soybean yield and chemical composition in response to phosphorus-potassium nutrition in Kashmir. *Agron. J.* **2012**, *104*, 1476–1484. [CrossRef]
59. Aulakh, M.S.; Pasricha, N.S.; Bahl, G.S. Phosphorus fertilizer response in an irrigated soybean–wheat production system on a subtropical, semiarid soil. *Field Crops Res.* **2003**, *80*, 99–109. [CrossRef]
60. Ugochukwu, U.E.; Oke, O.F. Effects of rate of phosphorus fertilizer on crop growth and yield in rice-soybean intercropping system. *Int. J. Biosci.* **2021**, *18*, 29–37.
61. Fituma, T.; Tana, T.; Alemneh, A.A. Response of soybean to Bradyrhizobium japonicum inoculation and phosphorus application under intercropping in the Central Rift Valley of Ethiopia. *South. Afr. J. Plant Soil* **2018**, *35*, 33–40. [CrossRef]
62. Dhage, S.J.; Patil, V.D.; Patange, M.J. Effect of various levels of phosphorus and sulphur on yield, plant nutrient content, uptake and availability of nutrients at harvest stages of soybean [*Glycine max* (L.)]. *Int. J. Curr. Microbiol. Appl. Sci.* **2014**, *3*, 833–844.
63. Abbasi, M.K.; Manzoor, M.; Tahir, M.M. Efficiency of Rhizobium inoculation and P fertilization in enhancing nodulation, seed yield, and phosphorus use efficiency by field grown soybean under hilly region of Rawalakot Azad Jammu and Kashmir, Pakistan. *J. Plant Nutr.* **2010**, *33*, 1080–1102. [CrossRef]
64. Abbasi, M.K.; Majeed, A.; Sadiq, A.; Khan, S.R. Application of Bradyrhizobium japonicum and phosphorus fertilization improved growth, yield and nodulation of soybean in the sub-humid hilly region of Azad Jammu and Kashmir, Pakistan. *Plant Prod. Sci.* **2008**, *11*, 368–376. [CrossRef]
65. Rehm, G.W. Response of Irrigated Soybeans to Rate and Placement of Fertilizer Phosphorus. *Soil Sci. Soc. Am. J.* **1986**, *50*, 1227–1230. [CrossRef]
66. Kolawole, G.O. Effect of phosphorus fertilizer application on the performance of maize/soybean intercrop in the southern Guinea savanna of Nigeria. *Arch. Agron. Soil Sci.* **2012**, *58*, 189–198. [CrossRef]

67. Ronner, E.; Franke, A.C.; Vanlauwe, B.; Dianda, M.; Edeh, E.; Ukem, B.; Bala, A.; Van Heerwaarden, J.; Giller, K.V. Understanding variability in soybean yield and response to P-fertilizer and rhizobium inoculants on farmers' fields in northern Nigeria. *F Crop Res.* **2016**, *186*, 133–145. [CrossRef]
68. Ahmed, S.; Raza, M.A.; Zhou, T.; Hussain, S.; Khalid, M.H.B.; Feng, L.; Wasaya, A.; Iqbal, N.; Ahmed, A.; Liu, W.; et al. Responses of Soybean Dry Matter Production, Phosphorus Accumulation, and Seed Yield to Sowing Time under Relay Intercropping with Maize. *Agronomy* **2018**, *8*, 282. [CrossRef]
69. Shewangizaw, B.; Kassie, K.; Assefa, S.; Feyisa, T. On farm verification of soil test-based phosphorus fertilizer recommendations for bread wheat (*Triticum aestivum* L.) on the vertisols of central highlands of Ethiopia. *Cogent Food Agric.* **2020**, *6*, 1807811. [CrossRef]
70. IndexMundi. Available online: www.indexmundi.com (accessed on 21 July 2022).
71. Van Rossum, G.; Drake, F.L. *Python/C Api Manual-Python 3*; CreateSpace: Scotts Valley, CA, USA, 2009.
72. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
73. Bomeisl, L.P.; Neill, C.; Porder, S.; Cerri, C.E.; Brando, P.M.; Roy, E.D. Tropical soybean yield response to reduced or zero phosphorus fertilization depends on soils. *Agrosyst. Geosci. Environ.* **2020**, *3*, e20113. [CrossRef]
74. Rheinheimer, D.S.; Anghinoni, I.; Kaminski, J. Depletion of inorganic phosphorus from different fractions caused by successive extraction with resin in different soils management. *Rev. Bras. Ciênc. Solo* **2000**, *24*, 345–354. [CrossRef]
75. Ferguson, R.B.; Shapiro, C.A.; Dobermann, A.R.; Wortmann, C.S. *Fertilizer Recommendation for Soybean*; NebGuide G859; University of Nebraska: Lincoln, NE, USA, 2006. Available online: http://www.ianrpubs.unl.edu/live/g859/build/g859.pdf (accessed on 28 August 2022).
76. Khojely, D.M.; Ibrahim, S.E.; Sapey, E.; Han, T. History, current status and prospects of soybean production and research in sub-Saharan Africa. *Crop J.* **2018**, *6*, 226–235. [CrossRef]
77. Masuda, T.; Goldsmith, P.D. World Soybean Production: Area Harvested, Yield, and Long-Term Projections. *Int. Food Agribus. Manag. Rev.* **2009**, *12*, 143–163.
78. Langemeier, M.; Zhou, L. International Benchmarks for Soybean Production. *Farmdoc Dly.* **2022**, *12*, 43.
79. Olsen, S.R.; Khasawneh, F.E. Use limitations of physical-chemical criteria for assessing the status of phosphorous in soils. In *The Role of Phosphorous in Agriculture*; Khasawneh, F.E., Sample, E.C., Kamprath, E.J., Eds.; American Society of Agronomy: Madison, WI, USA, 1980; pp. 361–410.
80. Holford, I.C.R. Greenhouse evaluation of four phosphate soil tests in relation to phosphate buffering and labile phosphate in soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 555–559. [CrossRef]
81. Sims, J.T.; Ellis, B.G. Adsorption and availability of phosphorous following the application of limestone to an acid, aluminous soil. *Soil Sci. Soc. Am. J.* **1983**, *47*, 888–893. [CrossRef]
82. Rehm, G.W.; Schmitt, M.A.; Lamb, J.; Eliason, R. *Fertilizing Soybeans in Minnesota*; University of Minnesota Extension Service: St Paul, MN, USA, 2001. Available online: www.extension.umn.edu/distribution/cropsystems/DC3813.html (accessed on 18 August 2022).
83. Sawyer, J.E.; Mallarino, A.P.; Killorn, R.; Barnhart, S.K. *A General Guide for Crop Nutrient and Limestone Recommendations in Iowa*; Iowa State University Extension and Outreach: Ames, IA, USA, 2008.
84. Anthony, P.; Malzer, G.; Sparrow, S.; Zhang, M. Soybean yield and quality in relation to soil properties. *Agron. J.* **2012**, *104*, 1443–1458. [CrossRef]
85. Franzen, D. *North Dakota Fertilizer Recommendation Tables and Equations*; SF882; Revised; NDSU Extension Service: Fargo, ND, USA, 2018; pp. 1–16. Available online: https://www.ag.ndsu.edu/publications/crops/north-dakota-fertilizer-recommendation-tables-and-equations/sf882.pdf (accessed on 23 February 2023).
86. Wortmann, C.S.; Krienke, B.T.; Ferguson, R.B.; Maharjan, B. Fertilizer Recommendations for Soybean. NebGuide G859. 2018. Available online: http://extension.unl.edu/publications (accessed on 12 August 2022).
87. Bray, R.H. A nutrient mobility concept of soil-plant relationships. *Soil Sci.* **1954**, *78*, 9–22. [CrossRef]
88. Morris, N.; Knight, S.; Philpott, H.; Blackwell, M. *Cost-Effective Phosphorus Management on UK Arable Farms Report on Work Package 2: Critical Levels of Soil*; Agriculture and Horticulture Development Board (AHDB): Stoneleigh Park, UK, 2017.
89. Gerwing, J.; Gelderman, R.; Clark, J. Fertilizer Recommendation Guide. 2019. Available online: https://extension.sdstate.edu/sites/default/files/2019-03/P-00039_0.pdf (accessed on 12 August 2022).
90. Center of Reference in Agriculture and Agri-Food of Quebec (CRAAQ). *Reference Guide in Fertilization*, 2nd ed.; Center of Reference in Agriculture and Agri-Food of Quebec (CRAAQ): Québec City, QC, Canada, 2010; 473p.
91. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef] [PubMed]
92. Frederick, J.R.; Camp, C.R.; Bauer, P.J. Drought-stress effects on branch and mainstem seed yield and yield components of determinate soybean. *Crop Sci.* **2001**, *41*, 759–763. [CrossRef]
93. Sadeghipour, O.; Abbasi, S. Soybean response to drought and seed inoculation. *World Appl. Sci. J.* **2012**, *17*, 55–60.
94. Suriyagoda, L.D.; Ryan, M.H.; Renton, M.; Lambers, H. Plant responses to limited moisture and phosphorus availability. *Adv. Agron.* **2014**, *124*, 143–200. [CrossRef]
95. Gupta, S.; Jain, N.; Chopade, A.; Bhonde, A. A Machine Learning Based Application for Agricultural Solutions. *arXiv* **2022**, arXiv:2204.11340.

96. Lima Neto, A.J.; Deus, J.A.L.; Rodrigues Filho, V.A.; Natale, W.; Parent, L.E. Nutrient diagnosis of fertigated "Prata" and "Cavendish" banana (*Musa* spp.) at plot-scale. *Plants* **2020**, *9*, 1467. [CrossRef]
97. Kyveryga, P.M.; Blackmer, A.M.; Morris, T.F. Disaggregating Model Bias and Variability when Calculating Economic Optimum Rates of Nitrogen Fertilization for Corn. *Agron. J.* **2007**, *99*, 1048–1056. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.