MULTI-SCIENCE
JOURNAL

ISSN 2359-6902

Vol. 9 (2)
2026

Agricultural Sciences

Human Sciences

Biological Sciences

Health Sciences

Environmental Sciences

Exact and Earth Sciences

Veterinary Sciences

Influence of soil physicochemical variables on soybean yield across different soil profiles

Walter Baida Garcia Coutinho [1*], Verônica Manhães SantClair [1], Paulo Roberto Cecon [1], Anderson Rodrigo da Silva [2], Wilhan Valasco dos Santos [2], Sebastião Martins Filho [1], Antônio Policarpo Souza Carneiro [1] and Ana Carolina Campana Nascimento [1]

[1] *Departamento de estatística, Universidade Federal de Viçosa (UFV), Viçosa, Minas Gerais, Brazil. Avenida Peter Henry Rolfs, Campus Universitário, Viçosa–MG, 36570-900, Brazil.*
[2] *Laboratório de estatística e geoprocessamento, Instituto Federal Goiano - Campus Urutaí. Rod. Geraldo Silva Nascimento, km 2.5, 75780-000, Urutaí - GO.*

* *Corresponding author. E-mail: walterbgc1@gmail.com*

ABSTRACT. Deep soil fertility management is decisive for the productive potential of soybean because it expands the volume of soil effectively explored by roots and reduces physical and chemical constraints to plant growth. This study aimed to quantify the relative contribution of soil physicochemical attributes, sampled from 0 to 200 cm depth, to the prediction of soybean yield in high-performance fields from the Brazilian Soybean Strategic Committee (CESB). A Random Forest classification model was fitted and evaluated using out-of-bag (OOB) error, class purity–based metrics, SHAP values for model interpretation, and partial dependence curves combined with a purity metric across yield classes. The model showed adequate predictive performance (OOB error ≈ 17.5%), and five predictors were consistently important across yield classes and depths: clay content, cation exchange capacity (CEC), phosphorus, pH and copper. We conclude that management decisions targeted at high yields should consider sufficiency levels by depth layer, integrating liming, gypsum application, fertilization and decompaction practices that maximize the volume of soil that can be effectively explored by roots and the extension of the ideally exploitable profile. The results reinforce the need for deep soil diagnosis, with sampling beyond the traditional 0–20 and 20–40 cm layers, and demonstrate the potential of machine-learning approaches to integrate large volumes of soil data and support more accurate, yield-oriented management recommendations.

Key words: random forest; deep soil profile; management depth; relative variable importance; soybean yield; machine learning.

DOI: https://doi.org/10.33837/msj.v9i2.1799

Received: March 5th, 2026 *Accepted: May 23th, 2026* *Online published: June 8th, 2026*
Associate editor: Herbert J. Dias

INTRODUCTION

Soybean yield is limited by edaphic constraints throughout the soil profile, while soil management on Brazilian farms is mostly restricted to depths of up to 40 cm. It is still not fully clarified which deep soil physicochemical attributes explain yield variability (Frene et al., 2024; Alves et al., 2025).

At the same time, modern cultivars demand more nutrients to sustain higher yield ceilings. Modern genotypes remove 15–23% (Filippi et al., 2021) and 11–45% (Esper Neto et al., 2021) more nutrients than most reference values described in official fertilizer recommendation systems (Bender et al., 2015). The importance of managing the soil profile is therefore crucial, especially for high nutrient-demanding crops, in which soil structure can be decisive for maximizing yield (Oliveira et al., 2024; Silva et al., 2025).

Studies such as those by Caires and Guimarães (2018) and Bossolani et al. (2021) offer more up-to-date approaches and help mitigate concerns about the use of high rates of soil amendments. Proper management of cover crops is essential to maximize water infiltration and lime percolation, increasing porosity through root channels and root decomposition at depth, thereby favoring nutrient incorporation in deeper layers (Crusciol et al., 2019; Silva et al., 2022; Bartosiewicz et al., 2025). Currently, the average soybean yield in Brazil is 3,180 kg ha^{-1} (CONAB, 2024), whereas the average productive potential could reach 5,460 kg ha^{-1} (Marin et al., 2022), which represents approximately 71% more than the current national mean.

Reports from the Brazilian Soybean Strategic Committee (CESB) document fields achieving yields of 5,400 to 8,400 kg ha^{-1}, and show that the main difference between these areas and the rest of the country lies in deep soil management. These fields present soils free of aluminum toxicity, with adequate Ca, Mg and P levels at depth, as well as nodules at 140 cm and roots reaching 200 cm depth (CESB champion 2022/2023 – 8,061.6 kg ha^{-1}) (CESB, 2023). This evidences that the larger the volume of soil explored by roots, the higher the yield, a principle that applies not only to soybean but to all crops.

Despite the accumulated evidence on the benefits of deep soil management, there is still a lack of information regarding the relative importance of deep soil attributes that can support targeted management strategies. Machine-learning (ML) models have been used with predictive purposes in several contexts, such as yield and climate forecasting, but it is still necessary to detail deep soil characteristics that explain within-field yield variability (von Bloh et al., 2023; Berveglieri et al., 2024). The objective of this study was to determine the relative importance of soil physical and chemical variables down to 200 cm depth for predicting soybean yield.

MATERIALS AND METHODS

Data collection

The soil profile data used in this study were obtained from publicly available champion case reports from the Brazilian Soybean Strategic Committee (CESB, 2024). These reports are audited by three independent and specialized organizations, ensuring the reliability and transparency of the results generated within the scope of the CESB Challenge.

The reports were downloaded from the CESB website, and all data were manually extracted with double-checking to ensure accuracy. The dataset covers the 2017/18 to 2023/24 growing seasons and comprises 34 soil profiles, each sampled at 11 depths (10, 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 cm), totaling 374 layers. Soybean yields ranged from 5,400 to 8,400 kg ha^{-1}. For analysis, the profiles were stratified into three yield classes: 5,400–6,300; 6,300–7,200; and 7,200–8,400 kg ha^{-1}, with 8, 17 and 9 profiles (equivalent to 88, 187 and 99 layers), respectively. These values also represent the number of observations considered in the database for each yield class.

The variables collected are described in Table 1. Variable selection was based on the availability of data for all evaluated depths. We chose to prioritize chemical and some physicochemical variables (such as clay content) because they are routinely determined in soil fertility analyses, which favors reproducibility and comparability with other studies. In addition, part of the effects of physical properties is indirectly captured by clay content, which strongly influences water retention, hydraulic conductivity and nutrient availability. Studies such as those by Gazolla-Neto et al. (2016) and Cox et al. (2003) also adopted clay as a representative texture variable, avoiding the redundant inclusion of other physical fractions when the goal was to identify key attributes related to yield.

Table 1. Variables used for soil profile analysis and for the predictive study of soybean yield.

Abbreviation	Description	Unit
Arg	Clay	%
V	Base saturation	%
pH	Hydrogen potential	$CaCl_2$
M.O	Organic matter	g dm^{-3}
CEC	Cation exchange capacity	mmolc dm^{-3}
K	Potassium	mmolc dm^{-3}
Ca	Calcium	mmolc dm^{-3}
Mg	Magnesium	mmolc dm^{-3}
Al	Aluminum	mmolc dm^{-3}
P	Phosphorus	mg dm^{-3}
S	Sulfur	mg dm^{-3}
B	Boron	mg dm^{-3}
Fe	Iron	mg dm^{-3}
Cu	Copper	mg dm^{-3}
Mn	Manganese	mg dm^{-3}
Zn	Zinc	mg dm^{-3}

Data analysis

To predict the three soybean yield classes, a Random Forest classification model was fitted using the variables in Table 1 as predictors. The data were split into training and testing sets in a 70% and 30% proportion, respectively, using the createDataPartition function from the *caret* package (Kuhn, 2008). Hyperparameter tuning was performed by 5-fold cross-validation. The final model was fitted with 500 trees and mtry = 9, which defines the number of predictor variables randomly sampled at each split. As a measure of lack of fit, we reported the out-of-bag (OOB) error.

Model performance was evaluated based on accuracy and the confusion matrix obtained from the classifications of the test set, using the confusion Matrix function from the *caret* package (Kuhn, 2008). We computed the overall accuracy with 95% confidence interval, sensitivity, specificity, class-wise balanced accuracy, the p-value for comparison with the no-information rate, and the misclassification error (%).

From the confusion matrix, the test misclassification error was calculated for each yield class, defined as the complement of the proportion of correctly classified samples within that class, according to:

Equation 1.

$$Error_{(test)} = 1 - \frac{Correct\ classifications\ per\ class}{Total\ observations\ per\ class}$$

This indicator expresses the fraction of misclassified samples relative to the total number of observations belonging to each class, allowing the quantification of the class-specific error rate.

The relative importance of the variables was assessed using two metrics: the mean decrease in Gini index and the mean SHapley Additive exPlanations (SHAP) value for each variable. Both metrics were converted to percentages to facilitate interpretation (Lundberg & Lee, 2017). While the Gini index provides a global measure of how much each predictor contributes to reducing node impurity across the forest, SHAP values provide a local and additive interpretation of the contribution of each predictor to individual model predictions.

Because the Random Forest model was fitted as a multiclass classifier, SHAP values were interpreted in relation to the predicted probability of each yield class (Lundberg & Lee, 2017). For observation *i*, the SHAP value of predictor *j* for yield class *k*, denoted by ij(k), is defined as the weighted average of the marginal contributions of j over all possible subsets of predictors S:

Equation 2.

$$\Phi_{ij}^{(K)} = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{Su\{j\}}^{(k)}(x_{i,Su\{j\}}) - f_S^{(k)}(x_{i,S}) \right]$$

Where *F* is the set of all predictors, *S* is a subset of predictors that does not include *j*, $f_s^{(k)}(x_{i,S})$ represents the expected model prediction for class *k* when only the predictors in *S* are considered and the remaining predictors are marginalized out, and the term in brackets represents the marginal contribution of predictor *j* when it is added to subset *S*. Thus, the prediction for class *k* can be decomposed as:

Equation 3.

$$f^{(k)}(x_i) = \Phi_0^{(k)} + \sum_{j=1}^{p} \Phi_{\mathrm{ij}}^{(\mathrm{k})}$$

Where Φ_0^k is the baseline prediction for class k, and $\Phi_{ij}^{(k)}$ represents the contribution of predictor *j* to shifting the prediction of observation *i* away from this baseline.

SHAP values are per-variable attributions that decompose the model prediction into the additive contributions of each predictor. These values were displayed in a beeswarm plot to show their directional effect, depicting the contribution of high and low values of each variable to the predicted probability of each class. Point color reflects the actual value of the variable (high/low) in the sampled data, allowing one to visualize simultaneously which variables influence the prediction and in which direction.

For variables showing agreement between Gini and SHAP metrics and importance values above the 6.25% cutoff (1/16), partial dependence plots (PDPs) were generated by yield class, using class purity as the response. Shaded bands representing the 10–90% quantiles highlight the sampled regions for each class, while the rug at the bottom represents the density of observed values.

Purity was used as a measure of confidence in the predictions, defined as:

Equation 4.

$$Purity(x) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{k} \hat{p}_{ik}\,(x)^2$$

Where $\hat{p}_{ik}(x)$ is the predicted probability for class *k* for observation *i* when the variable of interest is fixed at value *x*, while all other variables remain at their observed values. The term *x* therefore represents the specific value assumed by the variable under analysis in the PDP, varying along a grid of values within the observed domain. In these curves, only the variable of interest is changed, and the output corresponds to the mean purity associated with each value of *x*. All analyses were performed in R version 4.5.1 (R Core Team, 2025).

RESULTS AND DISCUSSION

The confusion matrix shows that the model achieved an accuracy of 86.49%, which is sufficient to capture a substantial portion of the data variation and to identify which variables have the greatest impact on the predictions. Sensitivity (true positive rate) ranged from 69 to 96%, whereas specificity (true negative rate) ranged from 83 to 98% across yield classes, and the out-of-bag (OOB) predictive error was 17.49% (Table 2).

The overall test error was 13.51%, indicating high predictive ability and good generalization of the model to the test set. The higher error rates observed in the upper yield classes (5.26, 14.29, and 17.24%) may be associated with overlapping soil physicochemical characteristics among groups and with the increasing complexity of fertility gradients as yield rises. These results reinforce the robustness of the Random Forest model for classifying productive environments, with performance consistent with the natural complexity of soil attributes that determine soybean yield.

Given that the final model included 16 predictor variables, the expected importance value for each of them, under an assumption of equal contribution, would be 1/16, i.e., 6.25%. Thus, any variable with relative importance below this threshold can be classified as having low relevance within the context of the fitted model. This approach provides an objective criterion to discriminate variables with marginal contribution to the predictive capacity of the model, avoiding overinterpretation of effects associated with attributes that exert little influence on yield.

Table 2. Training and test confusion matrices for soil profile (0–200 cm) classification using the Random Forest model.

Training set			
Observed yield class (kg ha^{-1})	5.400-6.300	6.300-7.200	7.200-8.400
5.400-6.300	45	15	2
6.300-7.200	2	122	7
7.200-8.400	2	18	50
Test set			
Observed yield class (kg ha^{-1})	5.400-6.300	6.300-7.200	7.200-8.400
5.400-6.300	18	1	0
6.300-7.200	4	54	5
7.200-8.400	4	1	24
Sensitivity	0,69	0,96	0,82
Specificity	0,98	0,83	0,93
Balanced Accuracy	0,84	0,90	0,88
Overall accuracy	0,8649 (IC: 0,787 – 0,922)		
Out-of-bag error	17,49 (%)		
p-value [Acc > NIR]	$1{,}28e^{-15}$		
Class error 5.400–6.300 (Test)	5,26 (%)		
Class error 6.300–7.200 (Test)	14,29 (%)		
Class error 7.200–8.400 (Test)	17,24 (%)		
Overall error (Test)	13,51 (%)		

Using this cutoff (6.25% or 1/16), both Gini and SHAP metrics highlighted the predictors that were most relevant for the model. Clay (Gini 13.29%; SHAP 14.44%), cation exchange capacity – CEC (11.89%; 18.41%), P (10.69%; 9.89%), pH (8.18%; 11.05%) and Cu (6.95%; 8.88%) exceeded the threshold in both measures, identifying them as variables that meaningfully affect the confidence of the predictions. The remaining variables showed importance values below 6.25% in at least one of the metrics and were therefore classified as having low relevance. The satisfactory performance of the model (out-of-bag error of 17.49%) provides a solid basis for trusting the set of predictor variables identified (Table 3).

The simultaneous selection of clay, CEC, P, pH, and Cu indicates that the separation among yield classes did not depend on a single isolated attribute, but rather on the integration of water and nutrient retention capacity, chemical availability, and favorable conditions for root development throughout the soil profile. This interpretation is consistent with studies showing that soybean yield responds to the interaction among chemical, physical, and biological soil attributes at different depths, and not only to the conditions of the surface layer (Müller et al., 2021; Barbosa et al., 2025).

In highly weathered tropical soils, this integration is particularly important because acidity, low P availability, and the restricted distribution of exchangeable bases at depth can limit root growth and reduce the soil volume effectively explored by plants (Moraes et al., 2023; Oliveira et al., 2024).

Table 3. Relative importance of soil variables (0–200 cm depth) for predicting soybean yield classes using the Random Forest model.

Predictor	**Gini**	**SHAP**	**Gini (%)**	**SHAP (%)**
Arg	21,78	0,06	13,29	14,44
CEC	19,49	0,07	11,89	18,41
P	17,52	0,04	10,69	9,89
pH	13,41	0,04	8,18	11,05
Cu	11,39	0,04	6,95	8,88
B	10,20	0,02	6,22	4,06
S	9,43	0,01	5,75	2,82
Mg	9,40	0,03	5,74	6,54
Zn	8,92	0,03	5,44	6,31
Fe	8,38	0,02	5,12	4,58
V	7,42	0,01	4,53	2,81
K	7,31	0,01	4,46	2,75
M.O.	6,75	0,01	4,12	2,45
Mn	5,92	0,01	3,61	2,08
Ca	5,16	0,01	3,15	2,25
Al	1,43	0,00	0,87	0,70
Total	163,886	0,4	100	100
Out-of-bag (%)			17,49%	

Clay stood out with high importance (13.29%; 14.44%) because it is directly related to soil water and nutrient retention (Table 3). Soils with higher clay and organic matter contents tend to have greater water-holding capacity and higher cation exchange capacity – CEC (11.89%; 18.41%), which allows greater retention of cations, favoring root development and nutrient uptake (Teixeira et al., 2017). This relationship was also highlighted by Vitantonio-Mazzini et al. (2020), who observed that the interaction between water availability and soil physicochemical characteristics was decisive for maximizing soybean yield.

This retention capacity, combined with a more uniform distribution of nutrients throughout the soil profile, promotes deeper root development. It is worth noting that the CEC of organic matter has a higher magnitude than that of clay in many tropical soils, since most of these soils are derived from kaolinite, which has low CEC (Raij, 2011).

Therefore, the joint importance of clay and CEC should be interpreted as an indication that high-yield environments depend not only on the chemical capacity to retain nutrients, but also on physical conditions that allow greater root exploration at depth, as deeper chemical correction can favor root growth and crop yield in highly weathered soils (Moraes et al., 2023).

Phosphorus showed high importance (10.69%; 9.89%), indicating that, even at depth, it exerts a direct influence on yield. This finding corroborates studies showing a strong association between available P and soybean yield, even below the surface layer (Esper Neto et al., 2021; Gazolla-Neto et al., 2016). In areas with higher P levels at depth, yield tends to be greater, possibly due to the larger soil volume explored by roots and the maintenance of P uptake during periods of water deficit. Sawchik and Mallarino (2008) further demonstrated that the spatial variability of P and K is closely related to field-level yield, which makes deep nutrient management even more relevant.

Soil pH, by influencing nutrient availability, also had a marked importance (8.18%; 11.05%) for soybean yield (Table 3). Malavolta (1979) describes in detail the interaction among nutrients at different pH levels, emphasizing that when pH is outside the optimum range for the crop, plant uptake is limited even if nutrients are present in the soil. In this study, pH variability proved important for discriminating yield levels, which is reflected in the high importance assigned to this variable.

The high relevance of Cu (6.95%; 8.88%) underscores the importance of micronutrients. Copper is critical for soybean physiological performance because it directly participates in photosynthesis and respiration, which ultimately affects yield components (Moreira & Moraes, 2019). According to Barik and Chandel (2001), copper applications of up to 5 kg ha^{-1} increase nodulation, leaf area, dry matter and yield, and doses around 2.5 kg ha^{-1} can also enhance P uptake.

The joint presence of P, pH, and Cu among the most important variables reinforces that high yield depends not only on the total amount of nutrients in the soil, but also on their effective availability in the root environment. Soil pH acts as a central regulator of the solubility, adsorption, and availability of several nutrients, although its effects depend simultaneously on soil properties and plant response (Barrow & Hartemink, 2023).

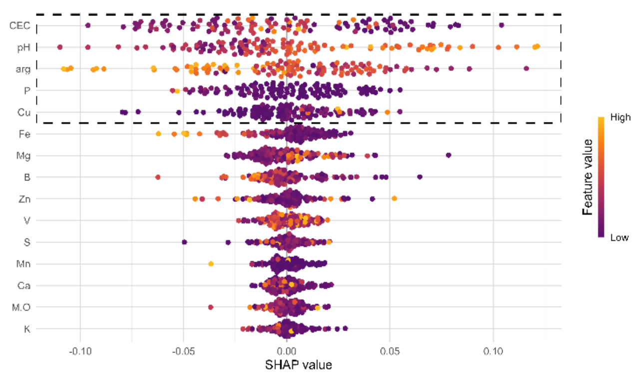
For P, this relationship is especially relevant in highly weathered tropical soils, where its availability is strongly controlled by adsorption and precipitation processes involving soil mineral constituents (Penn & Camberato, 2019; Hanyabui et al., 2020). In addition, the relevance of Cu is consistent with its role in soybean physiological processes and with its availability being conditioned by attributes such as organic matter and clay content (Moreira & Moraes, 2019). Thus, these attributes should be interpreted in an integrated manner, since changes in soil acidity can modify phosphorus availability, micronutrient availability, and nutrient uptake efficiency by plants.

Organic matter is a key component of soil, influencing structure, water retention and CEC. Its decomposition stimulates microbial activity and gradually releases nutrients, increasing soil fertility over time, including phosphorus, for which it can account for 20–80% of the total pool (Rosa et al., 2018; Batista et al., 2018). In addition, Cu dynamics is strongly modulated by organic matter through the action of humic and fulvic acids (Moreira & Moraes, 2019). In the present study, however, organic matter showed low importance (4.12%; 2.45%) because it is largely concentrated in the upper soil layers; when the entire 0–200 cm profile is considered, its effect is diluted and overshadowed by variables that are present and vary throughout the whole profile (Table 3).

The spread of the predictors in the beeswarm plot reflects the magnitude of their contribution to the model predictions and highlights the role of CEC, pH, clay, P and Cu (Figure 1) (Lundberg & Lee, 2017). The color gradient suggests biologically meaningful patterns: higher pH values (yellow points) tend to shift contributions towards positive regions, which is consistent with reduced acidity and improved availability of nutrients such as P (Havlin et al., 2014; Weil & Brady, 2017). Another clear example is clay content, where intermediate levels contribute positively, whereas very high levels tend to contribute negatively due to their effect on structural properties such as reduced macroporosity, aeration and infiltration, greater P sorption and micronutrient unavailability under heavy liming (Sposito, 2008; Fageria, 2009).

For P and Cu, a bilateral effect is observed, suggesting an optimal response within intermediate ranges. The remaining predictors, within the sampling range of this study, were not sufficiently informative to improve the classification of soil profiles.

Figure 1. SHAP values of soil variables in the model predictions.



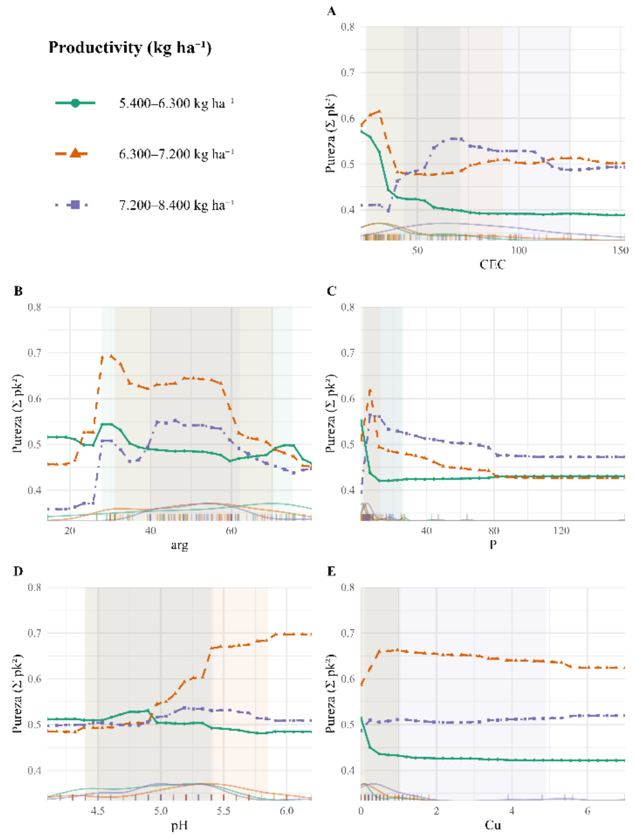
To analyze how soil attributes affect the confidence of yield-class predictions, partial dependence plots (PDPs) were generated for the mean purity of the predicted probabilities as a function of CEC, clay, P, pH and Cu (Figure 2A–E). High purity values indicate that the probability mass is concentrated in a single class, whereas lower values reflect greater uncertainty in classification.

Low CEC values increase purity for the low and medium yield classes, while the highest yield class is associated with intermediate-to-high CEC values, reaching a stable plateau at the upper end of the range (Figure 2A).

The purity of the lowest yield class tends to stabilize at clay contents above approximately 35%, indicating that higher clay levels are associated with a more favorable environment for medium yield levels, up to a saturation point at which purity decreases for all classes (Figure 2B).

For P, a rapid increase in purity is observed for the medium and high yield classes at low to moderate concentrations, followed by a plateau with a slight downward trend. The lowest yield class shows little gain in purity, suggesting that correcting P from deficient to adequate levels is decisive for achieving higher yields (Figure 2C). As pH is increased towards the optimum range, the purity of the medium and high yield classes also increases (Figure 2D).

Figure 2. Partial effects of soil variables on class probability purity for soybean yield classes.



At low Cu levels, the separation between yield classes is small; as Cu increases, the medium yield class shows higher purity, whereas the high yield class maintains an intermediate, more stable level (Figure 2E).

Taken together, the PDPs provide a practical guide for improving high-yield systems: (i) remove base-related chemical limitations (raise pH and consequently mitigate Al^{3+} toxicity); (ii) correct acute P and Cu deficiencies up to adequate ranges, considering deep management and profile building; and (iii) preserve or recover soil structure through practices such as no-tillage and aggregate formation, so that high clay content does not become a physical constraint (Minato et al., 2023; Oliveira et al., 2024).

The occurrence of plateaus for CEC, P and Cu reinforces the concept of sufficiency levels and efficient input allocation, whereby doses above the adequate range do not increase the probability of belonging to the highest yield classes, whereas a corrected profile expands the exploitable soil volume for water and nutrients.

This plateau behavior also indicates that the results should not be interpreted as a recommendation for indiscriminate increases in fertilizer or soil amendment rates, but rather as evidence that high-yield systems require the balanced correction of multiple limitations throughout the soil profile, especially because chemical improvement at depth can favor root growth and the use of water and nutrients in subsurface layers (Moraes et al., 2023; Oliveira et al., 2024).

This indicates that management strategies should consider not only the tilled layer but also the deeper profile, enabling high nutrient levels at different depths while avoiding toxicity and improving nutrient use efficiency, especially in terms of P availability and the maintenance of soil physical structure and water-holding capacity. Such an approach is supported by evidence that vertical and horizontal spatial variability directly influences input use efficiency and yield stability (Esper Neto et al., 2021; Cox et al., 2003).

CONCLUSION

Over the 0–200 cm profile, clay, CEC, P, pH and Cu stood out as the most consistent predictors of soybean yield. The largest increases in the probability of belonging to the highest yield classes were observed in soils with clay contents between 30 and 60%, CEC between 40 and 110 mmol dm^{-3}, available P above 20 mg dm^{-3}, pH between 5.3 and 6.2, and Cu between 0.5 and 1 mg dm^{-3}. Outside these ranges, a gradual reduction in prediction purity was observed, resulting in a higher probability of allocation to lower yield classes. These findings show that improving the chemical and structural conditions of the deep soil profile is directly associated with shifting environments towards higher yield levels.

ACKNOWLEDGMENTS

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors also thank the reviewers and editors for their valuable suggestions.

REFERENCES

Alves, L. A., Fontoura, S. M. V., Tiecher, T., Pesini, G., Flores, J. P. M., Filippi, D., Moraes, R. P. de Pias, O. H. de C., & Bayer, C. (2025). Long-term soil acidity dynamics and crop yield response to phosphogypsum and limestone with different reactivities in a no-till Oxisol. *European Journal of Agronomy, 171*, 127803. DOI: https://doi.org/10.1016/j.eja.2025.127803

Barik, K. C., & Chandel, A. S. (2001). Effect of copper fertilization on plant growth, seed yield, copper and phosphorus uptake in soybean (*Glycine max*) and their residual availability in Mollisol. *Indian Journal of Agronomy, 46*(2), 319–326. DOI: https://doi.org/10.59797/ija.v46i2.3266

Barrow, N. J., & Hartemink, A. E. (2023). The effects of pH on nutrient availability depend on both soils and plants. DOI: https://doi.org/10.1007/s11104-023-05960-5

Bartosiewicz, B., Wawer, R., Poręba, L., & Siebielec, G. (2025). The impact of cover crops on soil quality. *Current Agronomy, 54*(2), 118–128. DOI: https://doi.org/10.2478/cag-2025-0010

Barbosa, J. Z., Motta, A. C. V., Poggere, G., & Prior, S. A. (2025). Winners of the Brazilian soybean yield contest: Climatic, soil, management, and economic factors. *Oil Crop Science*, 10(4), 294–302. DOI: https://doi.org/10.1016/j.ocsci.2025.07.003

Batista, M. A., Inoue, T. T., Esper Neto, M., & Muniz, A. S. (2018). Princípios de fertilidade do solo, adubação e nutrição mineral. In J. U. T. Brandão Filho, P. S. L. Freitas, L. O. S. Berian, & R. Goto (Eds.), *Hortaliças-fruto* (pp. 113–162). EDUEM. DOI: https://doi.org/10.7476/9786586383010.0006

Bender, R. R., Haegele, J. W., & Below, F. E. (2015). Nutrient uptake, partitioning, and remobilization in modern soybean varieties. *Agronomy Journal, 107*(2), 563–573. DOI: https://doi.org/10.2134/agronj14.0435

Berveglieri, A., Imai, N. N., Watanabe, F. S. Y., Tommaselli, A. M. G., Ederli, G. M. P., Araújo, F. F. de, Lupatini, G. C., & Honkavaara, E. (2024). Remote prediction of soybean yield using UAV-based hyperspectral imaging and machine learning models. *AgriEngineering, 6*(3), 3242–3260. DOI: https://doi.org/10.3390/agriengineering6030185

Bossolani, J. W., Crusciol, C. A. C., Portugal, J. R., Moretti, L. G., Garcia, A., Rodrigues, V. A., Fonseca, M. de C. da, Bernart, L., Vilela, R. G., Mendonça, L. P., & Reis, A. R. dos. (2021). Long-term liming improves soil fertility and soybean root growth, reflecting improvements in leaf gas exchange and grain yield. *European Journal of Agronomy, 128*, 126308. DOI: https://doi.org/10.1016/j.eja.2021.126308

Caires, E. F., & Guimarães, A. M. (2018). A novel phosphogypsum application recommendation method under continuous no-till management in Brazil. *Agronomy Journal, 110*(5), 1987–1995. DOI: https://doi.org/10.2134/agronj2017.11.0642

Comitê Estratégico Soja Brasil. (2024). Desafio Nacional de Máxima Produtividade de Soja. Disponível em: URL: <https://www.cesbrasil.org.br/>

Companhia Nacional de Abastecimento. (2024). Acompanhamento da safra brasileira de grãos (Vol. 11, Safra 2023/24, No. 11, Décimo primeiro levantamento). CONAB. URL: https://www.gov.br/conab

Cox, M. S., Gerard, P. D., Wardlaw, M. C., & Abshire, M. J. (2003). Variability of selected soil properties and their relationships with soybean yield. *Soil Science Society of America Journal, 67*(4), 1296–1302. DOI: https://doi.org/10.2136/sssaj2003.1296

Crusciol, C. A. C., Marques, R. R., Carmeis Filho, A. C. A., Soratto, R. P., Costa, C. H. M. da, Ferrari Neto, J., Castro, G. S. A., Pariz, C. M., Castilhos, A. M., & Franzluebbers, A. J. (2019). Lime and gypsum combination improves crop and forage yields and estimated meat production and revenue in a variable charge tropical soil. *Nutrient Cycling in Agroecosystems, 115*(3), 417–432. DOI: https://doi.org/10.1007/s10705-019-10017-0

Esper Neto, M., Lara, L. M., Oliveira, S. M. de, Santos, R. F., Braccini, A. L., Inoue, T. T., & Batista, M. A. (2021). Nutrient removal by grain in modern soybean varieties. *Frontiers in Plant Science, 12*, 615019. DOI: https://doi.org/10.3389/fpls.2021.615019

Fageria, N. K. (2009). *The use of nutrients in crop plants*. CRC Press.

Filippi, D., Denardin, L. G. de O., Ambrosini, V. G., Alves, L. A., Flores, J. P. M., Martins, A. P., Pias, O. H. de C., & Tiecher, T. (2021). Concentration and removal of macronutrients by soybean seeds over 45 years in Brazil: A meta-analysis. *Revista Brasileira de Ciência do Solo, 45*, e0200186. DOI: https://doi.org/10.36783/18069657rbcs20200186

Frene, J. P., Pandey, B. K., & Castrillo, G. (2024). Under pressure: Elucidating soil compaction and its effect on soil functions. *Plant and Soil, 502*, 267–278. DOI: https://doi.org/10.1007/s11104-024-06573-2

Gazolla-Neto, A., Fernandes, M. C., Vergara, R. O., Gadotti, G. I., & Villela, F. A. (2016). Spatial distribution of the chemical properties of the soil and of soybean yield in the field. *Revista Ciência Agronômica, 47*(2), 325–333. DOI: https://doi.org/10.5935/1806-6690.20160038

Hanyabui, E., Apori, S. O., Frimpong, K. A., Atiah, K., Abindaw, T., Ali, M., Asiamah, J. Y., & Byalebeka, J. (2020). Phosphorus sorption in tropical soils. *AIMS Agriculture and Food*, 5(4), 599–616. DOI: https://doi.org/10.3934/agrfood.2020.4.599

Havlin, J. L., Tisdale, S. L., Nelson, W. L., & Beaton, J. D. (2014). *Soil fertility and fertilization: An introduction to nutrient management* (8th ed.). Pearson Prentice Hall.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. DOI: https://doi.org/10.18637/jss.v028.i05

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems* (Vol. 30). Curran Associates. DOI: https://doi.org/10.48550/arXiv.1705.07874

Malavolta, E. (1979). *ABC da adubação*. (4th ed.). Editora Agronômica CERES.

Marin, F. R., Zanon, A. J., Monzon, J. P., Andrade, J. F., Silva, E. H. F. M., Richter, G. L., Antolin, L. A. S., Ribeiro, B. S. M. R., Ribas, G. G., Battisti, R., Heinemann, A. B., & Grassini, P. (2022). Protecting the Amazon forest and reducing global warming via agricultural intensification. *Nature Sustainability, 5*(12), 1018–1026. DOI: https://doi.org/10.1038/s41893-022-00968-8

Minato, E. A., Brignoli, F. M., Esper Neto, M., Besen, M. R., Cassim, B. M. A. R., Lima, R. S., Tormena, C. A., Inoue, T. T., & Batista, M. A. (2023). Lime and gypsum application to low-acidity soils: Changes in soil chemical properties, residual lime content and crop agronomic performance. *Soil and Tillage Research, 234*, 105860. DOI: https://doi.org/10.1016/j.still.2023.105860

Moreira, A., & Moraes, L. A. C. (2019). Soybean response to copper applied to two soils with different levels of organic matter and clay. *Journal of Plant Nutrition, 42*(16), 1857–1867. DOI: https://doi.org/10.1080/01904167.2019.1655039

Moraes, F. A. de, Moreira, S. G., Peixoto, D. S., Silva, J. C. R., Macedo, J. R., Silva, M. M., Silva, B. M., Sanchez, P. A., & Nunes, M. R. (2023). Lime incorporation up to 40 cm deep increases root growth and crop yield in highly weathered tropical soils. *European Journal of Agronomy*, 144, 126763. DOI: https://doi.org/10.1016/j.eja.2023.126763

Müller, M., Schneider, J.R., Klein, V.A., Silva, E., Junior, J.P.S., Souza, A.M., & Chavarria, G. (2021). Soybean root growth in response to chemical, physical, and biological soil variations. *Frontiers in Plant Science*, 12, 602569. DOI: https://doi.org/10.3389/fpls.2021.602569

Oliveira, D. B. de, Lacerda, J. J. de J., Cavalcante, A. P., Bezerra, K. G., Silva, A. P. M. da, Miranda, A. C. G., Rambo, T. P., Maschio, R., Andrade, H. A. F. de, Costa, P. M., Sousa, C. A. F. de, Oliveira Júnior, J. O. L., Sagrilo, E., & Souza, H. A. de. (2024). Lime and gypsum rates effects in new soybean areas in the Cerrado of Matopiba, Brazil. *Agriculture, 14*(7), 1034. DOI: https://doi.org/10.3390/agriculture14071034

Penn, C. J., & Camberato, J. J. (2019). A critical review on soil chemical processes that control how soil pH affects phosphorus availability to plants. *Agriculture*, 9(6), 120. DOI: https://doi.org/10.3390/agriculture9060120

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. URL: https://www.R-project.org/

Raij, B. van. (2011). *Fertilidade do solo e manejo dos nutrientes*. International Plant Nutrition Institute.

Rosa, S. F. da, Reinert, D. J., Reichert, J. M., Fleig, F. D., Rodrigues, M. F., & Gelain, N. S. (2018). Propriedades físicas e químicas de um Argissolo sob cultivo de *Eucalyptus dunnii* Maiden no Pampa Gaúcho. *Ciência Florestal, 28*(2), 580–590. DOI: https://doi.org/10.5902/1980509832040

Sawchik, J., & Mallarino, A. P. (2008). Variability of soil properties, early phosphorus and potassium uptake, and incidence of soybean sudden death syndrome. *Agronomy Journal, 100*(5), 1450–1462. DOI: https://doi.org/10.2134/agronj2007.0303

Silva, G. F. da, Luperini, B. C. O., Barcelos, J. P. de Q., Putti, F. F., Mooney, S. J., & Calonego, J. C. (2025). Mechanical chiseling versus root bio-tillage on soil physical quality and soybean yield in a long-term no-till system. *Agronomy, 15*(5), 1249. DOI: https://doi.org/10.3390/agronomy15051249

Silva, J. M., Almeida, C. X. de, Pena, L. K., Jorge, R. F., Silva, L. R. da, & Duarte, I. R. G. (2022). Estimativa da macroporosidade e microporosidade em função de sistemas de manejo e plantas de cobertura em Latossolo Vermelho cultivado com soja. *Research, Society and Development, 11*(3), e54411326833. DOI: https://doi.org/10.33448/rsd-v11i3.26833

Sposito, G. (2008). *The chemistry of soils* (2nd ed.). Oxford University Press.

Teixeira, P. C., Donagemma, G. K., Fontana, A., & Teixeira, W. G. (2017). *Manual de métodos de análise de solo* (3rd ed.). Embrapa.

Vitantonio-Mazzini, L. N., Gómez, D., Gambin, B. L., Di Mauro, G., Iglesias, R., Costanzi, J., Jobbágy, E. G., & Borrás, L. (2020). Sowing date, genotype choice, and water environment control soybean yields in central Argentina. *Crop Science, 61*(1), 715–728. DOI: https://doi.org/10.1002/csc2.20315

Von Bloh, M., Nóia Júnior, R. de S., Wangerpohl, X., Saltık, A. O., Haller, V., & Kaiser, L. (2023). Machine learning for soybean yield forecasting in Brazil. *Agricultural and Forest Meteorology, 341*, 109670. DOI: https://doi.org/10.1016/j.agrformet.2023.109670

Weil, R. R., & Brady, N. C. (2017). *The nature and properties of soils* (15th ed.). Pearson.

To cite this paper, use:

Coutinho, W.B.D., SantClair, V.M., Cecon, P.R., Silva, A.R., Santos, W.V., Martins Filho, S., Carneiro, A.P.S., & Nascimento, A.C.C. (2026). Influence of soil physicochemical variables on soybean yield across different soil profiles. *Multi-Science Journal*, 9(2): 1-8. DOI: https://doi.org/10.33837/msj.v9i2.1799