Publication status: This preprint has been published elsewhere.
DOI of the published preprint: https://doi.org/10.71252/2177-34912026730004

Gaussian Process Regression as an Alternative to Kriging and SVM for Spatial Yield Prediction

Vinicius Rofatto, Samuel Philippe, George Deroco Martins

https://doi.org/10.1590/SciELOPreprints.11312

Submitted on: 2025-02-19
Posted on: 2025-02-27 (version 1)
(YYYY-MM-DD)

Gaussian Process Regression as an Alternative to Kriging and SVM for Spatial Yield Prediction

Samuel Philippe[1] (*); Vinicius Francisco Rofatto[1]; George Deroco Martins[1]

[1] Universidade Federal de Uberlândia, Programa de Pós-Graduação em Agricultura e Informações Geoespaciais, Monte Carmelo, MG, Brazil.
samuelphilippe89@gmail.com; vfrofatto@gmail.com; deroco@ufu.br
Corresponding author (*): samuelphilippe89@gmail.com
ORCID:
Corresponding author (*): https://orcid.org/0009-0001-3020-4595
Vinicius Francisco Rofatto: https://orcid.org/0000-0003-1453-7530
George Deroco Martins: https://orcid.org/0000-0001-9738-7325

ABSTRACT

Detecting spatial yield variability is essential for precision agriculture, as it reduces environmental impact and improves economic returns. This study evaluates Gaussian Process Regression (GPR), Ordinary Kriging (OK), and Support Vector Machine (SVM) under different sampling densities. GPR and OK perform similarly, with GPR showing a slight advantage in low-sampling conditions. With 322 samples, GPR achieves higher accuracy (RMSE = 0.64 t/ha, R^2 = 0.68) than OK (RMSE = 0.72 t/ha, R^2 = 0.60), while SVM performs worse (RMSE = 0.76 t/ha, R^2 = 0.55). Regardless of sample size, SVM-generated maps exhibit a smoothing effect, reducing sensitivity to local variations. OK remains effective but is more sensitive to sample density due to its reliance on the semivariogram model and the assumption of isotropy. These findings highlight GPR as a robust method for spatial yield prediction, particularly in sparse data conditions. The study was conducted in Patos de Minas, Brazil, using 795 georeferenced soybean yield samples over 3.7 hectares. From a practical perspective, GPR and OK remain strong candidates for yield interpolation, reinforcing the importance of model selection based on data availability and spatial variability.

Keywords: modeling; geostatistics; machine learning; interpolation; spatial analysis; predictive analysis.

INTRODUCTION

The accuracy of the spatial interpolation method used to generate soybean yield distribution maps directly influences their reliability. Methods such as Ordinary Kriging (OK) and Support Vector Machine (SVM) provide different approaches to spatial data interpolation, each with specific data requirements and analytical complexity.

OK remains a widely used geostatistical approach that models spatial correlations through a semivariogram, based on the assumption that nearby points exhibit greater similarity than distant ones. While it provides unbiased estimates and allows for uncertainty quantification, its performance depends on a well-defined spatial structure and tends to decline with sparse data or when stationarity assumptions are violated [1]. Alternatively, machine learning methods, such as SVM, can capture complex, nonlinear spatial relationships without requiring a predefined statistical model. However, their effectiveness depends on careful parameter tuning and the availability of a sufficiently large dataset, as performance tends to deteriorate when sample density is low [2]. Recent studies have compared geostatistical and machine learning approaches for spatial predictions in precision agriculture, emphasizing the importance of selecting appropriate interpolation methods [3].

Gaussian Process Regression (GPR) represents a promising alternative for crop yield modeling, offering a nonparametric, kernel-based probabilistic framework that eliminates the need to assume a predefined functional relationship between variables. GPR has been extensively applied in agriculture to model nonstationary and nonlinear relationships. For example, Campos-Taberner et al. [4] employed GPR to estimate the Leaf Area Index

(LAI) from smartphone images, demonstrating its potential for precision agriculture. Similarly, Martínez-Ferrer et al. [5] applied GPR for crop yield estimation and interpretability, integrating multisensor satellite observations and meteorological data to represent complex spatial and temporal dynamics. Additionally, Alebele et al. [6] explored Gaussian Kernel Regression to estimate crop yield from combined optical and Synthetic Aperture Radar (SAR) imagery, reinforcing its effectiveness in agricultural monitoring. These studies highlight GPR's capacity to handle sparse sampling scenarios and enhance yield prediction accuracy, positioning it as a robust tool for spatial interpolation in precision agriculture.

This study investigates the use of GPR to estimate and analyze the spatial variability of soybean yield. As a crop of significant global economic importance, soybean production requires an understanding of yield spatial variability to optimize input allocation and minimize the impacts of environmental and physiological factors on field performance [7]. Soybean yield depends on a combination of biotic and abiotic factors that directly influence its spatial distribution, posing challenges for accurate modeling.

To address these challenges, this study evaluates the accuracy of the GPR method under different sample sizes to determine the extent to which a reduction in sample density affects the quality of spatial estimates. Additionally, the performance of GPR is compared with SVM and OK, both implemented in the Smart-Map plugin integrated into QGIS.

MATERIAL AND METHODS

Study Area and Available Dataset

The data for this study were collected in the municipality of Patos de Minas, specifically in the Santana de Patos neighborhood (Figure 1). The study area is located at 18º 51' 32"S latitude and 46º 29' 49"W longitude (WGS84), with an average elevation of

approximately 832 m. The region's climate falls under the "Aw" category in Köppen's climate classification system, which corresponds to a humid subtropical or mild temperate climate. Additionally, the average temperature is 21.8 ºC, and the annual average precipitation is 1296 mm [8].

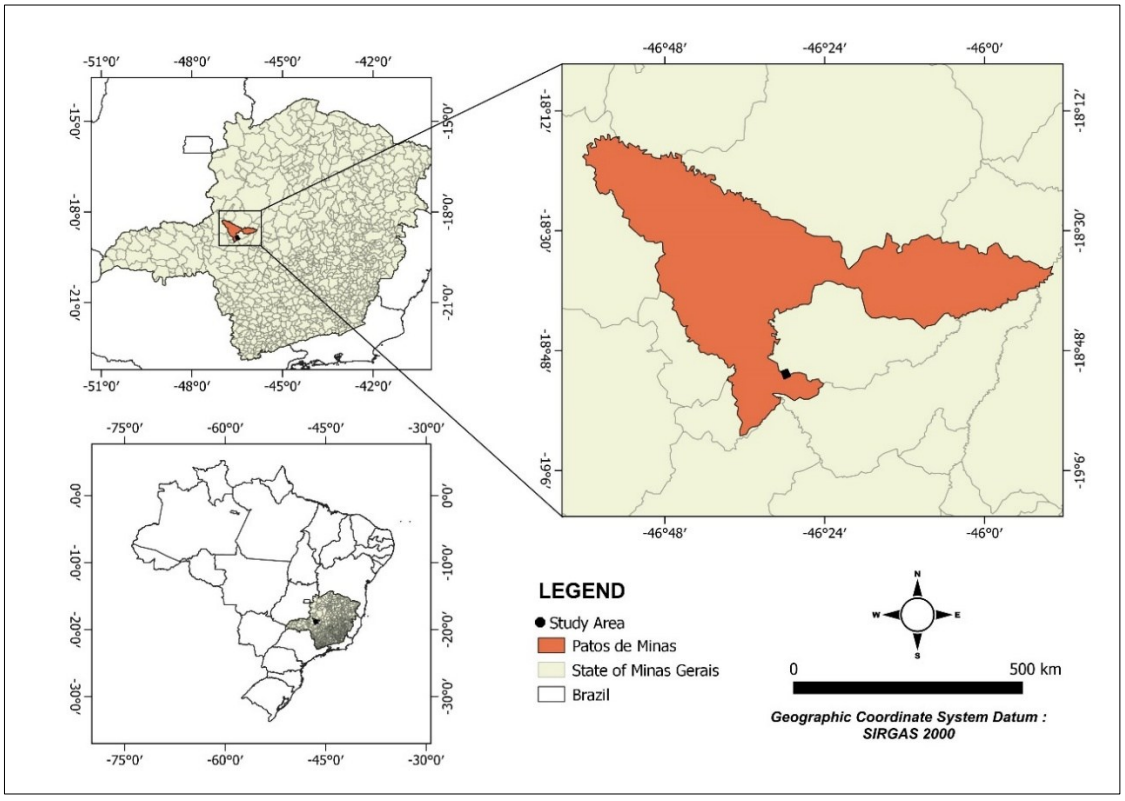

Figure 1: Study area

The study area covers approximately 3.7 hectares. Soybean seeds were planted in August 2022 and harvested in November 2022. Immediately after harvesting, the data were extracted from the grain harvester in ".csv" format and later converted into a shapefile format for processing.

This study utilized 795 sampling points, characterized by yield mass, measured in tons per hectare (t/ha), and spatial location in Cartesian coordinates within the UTM projection system (E, N) – 23S zone. The spatial distribution of the samples is shown in Figure 2, with coordinates standardized using the z-score method to ensure a mean of 0 and a standard deviation of 1. The highest yield values are observed in the central region of the

field (red circle), while the edges exhibit lower productivity (blue circle). This phenomenon is attributed to the border effect, which is common in cultivated areas where plants near the edges are more exposed to external factors such as microclimatic variations and pollutant contamination.

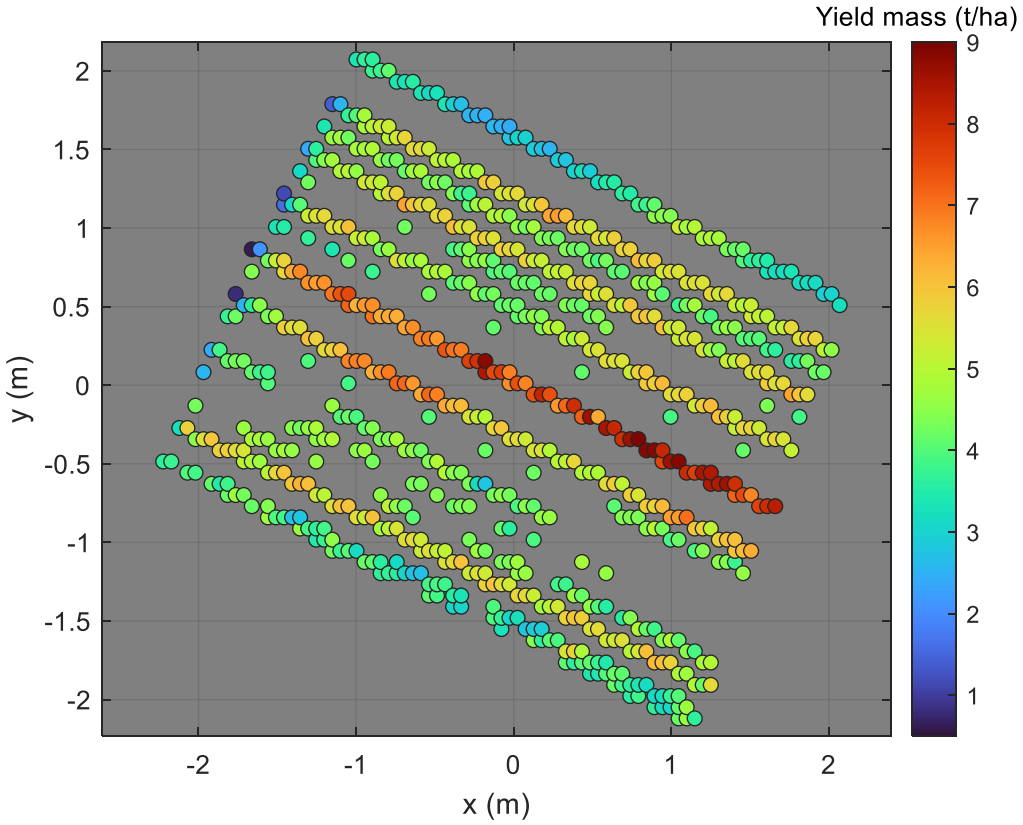

Figure 2: Spatial distribution of sampling points with yield variability in the field.

Gaussian Process Model for Soybeans Yield Estimation

Gaussian Process Regression (GPR) was employed to model and predict soybean yield based on spatial data. In this approach, the input variables consisted of the cartesian coordinates (x, y) representing the spatial location of each sampling point, while the output variable corresponded to yield mass (t/ha). The model was trained using the available sample data and subsequently applied to generate an interpolated yield map.

GPR is a nonparametric Bayesian regression method that defines a distribution over possible functions that fit the observed data. Given a set of training points $\boldsymbol{X} = \{(x_i, y_i)\}$

and corresponding yield values $\boldsymbol{Y} = \{y_i\}$, GPR estimates the yield at any unobserved location $(\bar{x}, \bar{y})$ by assuming that the observed values follow a multivariate Gaussian distribution. The model's behavior is determined by a covariance function (kernel), which defines the similarity between neighboring points.

In this study, the Rational Quadratic Kernel with a separate length scale per predictor was chosen due to its ability to handle multi-scale variations in spatial data, making it well-suited for yield prediction in heterogeneous fields. This approach has been shown to improve model generalization in environmental data applications by capturing both global and local variations in spatial relationships [9].

The Rational Quadratic Kernel is given by:

$$k(x, x') = \sigma^2 \left(1 + \frac{|x - x'|^2}{2\alpha\ell^2}\right)^{-\alpha} \quad (1)$$

where σ^2 is the variance parameter; ℓ is the length-scale parameter, adapted separately for each predictor (x, y); α is the scale-mixing parameter, controlling how much local variation is modeled; and $\boldsymbol{x} = (x, y)$ are spatial locations. More details about GPR can be found in Rasmussen and Williams, 2006 [10].

To assess generalization performance, k-fold cross-validation was conducted with $k = 5$, following the same procedure adopted in Ordinary Kriging (OK) and Support Vector Machine (SVM) in QGIS [11].

Experimental Setup

To ensure an unbiased evaluation, the dataset was randomly split into two subsets, namely the internal validation set in which 70% (557 sample points) of the data was used for k-fold cross-validation; and the external validation set which consisted of 30% (238 sample

points) of the original dataset, excluded from the entire k-fold cross-validation process and only used for final performance assessment. The external validation set was used to simulate real operational conditions, assessing the model's ability to generalize when new data becomes available. This realistic evaluation provides a more reliable estimate of the model's predictive performance beyond the k-fold cross-validation process [12].

Under the internal validation experiment, the original dataset used for training and cross- (557 sample points) was also progressively reduced to 90% (501), 80% (446), 60% (334), 40% (223), and 20% (112), as can be seen in Figure (3). For each reduced training set, k-fold cross-validation (k = 5) was applied to optimize the model and select the best hyperparameters.

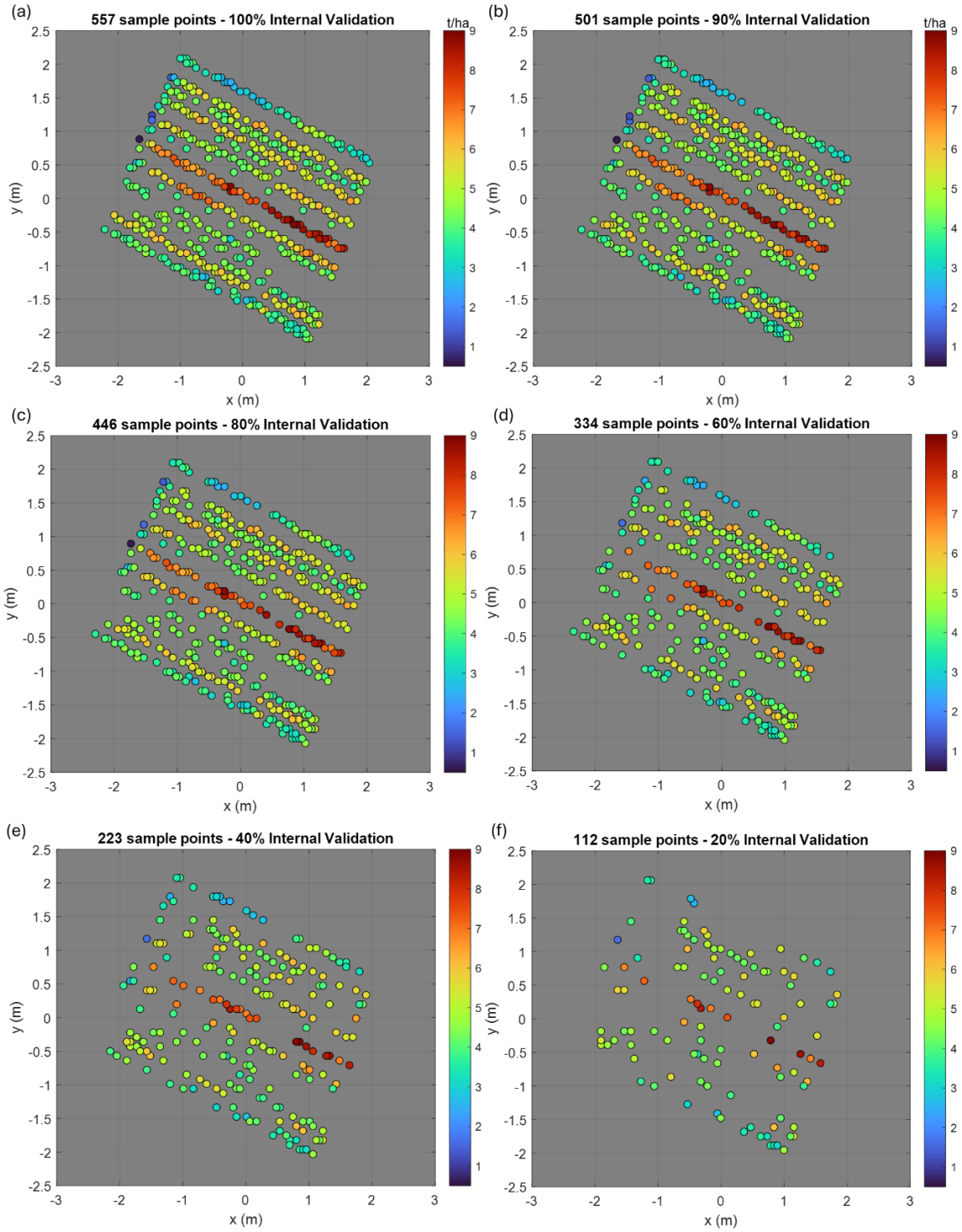

Figure 3: Progressive reduction of the training dataset used for GPR modeling: (a) 100% (557 points), (b) 90% (501 points), (c) 80% (446 points), (d) 60% (334 points), (e) 40% (223 points), and (f) 20% (112 points). The color scale represents soybean yield (t/ha), illustrating the point-by-point of yield across different sampling densities.

Once the models were selected during internal validation for each reduced sample scenario in Figure 3, they were evaluated on a fixed test set (238 out-of-sample points – external validation dataset), which remained unchanged throughout the entire process (Figure 4). This approach enables a systematic assessment of the impact of sample size reduction on GPR performance while ensuring an unbiased evaluation under real-world conditions, where new data become available. This step aimed at analyzing the model's generalization capability under different sampling constraints, particularly in scenarios with limited data availability, a critical challenge in spatial interpolation and predictive modeling [13].

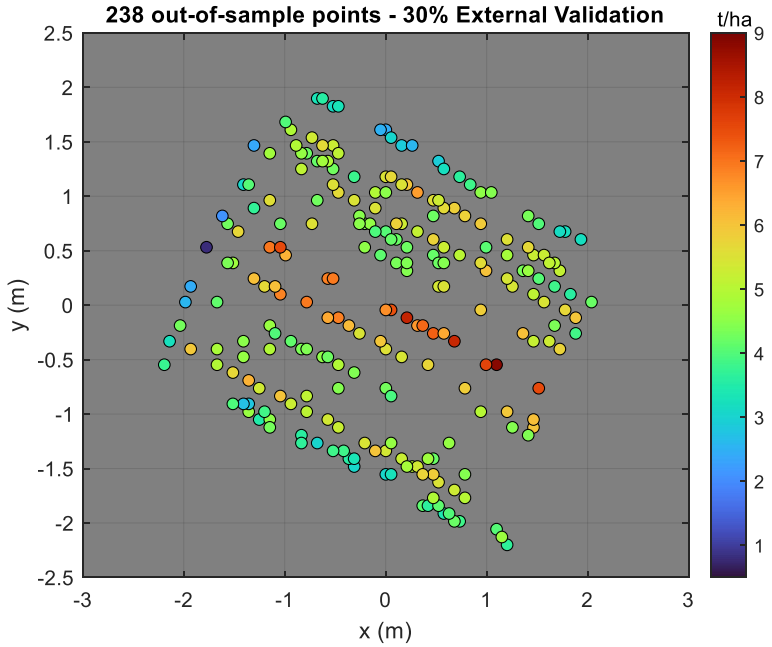

Figure 4: Spatial distribution of the 238 out-of-sample points used for external validation.

For each scenario presented in Figure 3, the corresponding GPR-fitted model was applied to a regular spatial grid covering the study area, enabling the interpolation of yield values at unsampled locations. The predicted yield values were then visualized as an interpolated yield map, facilitating the analysis of spatial productivity patterns and assessing the impact of different training sample sizes on model performance.

Subsequently, a second experiment evaluated GPR performance in comparison to OK and SVM, both implemented using the Smart-Map plugin in QGIS. In that case, two datasets were analyzed: one with 768 points and another with 322 points, which resulted from an outlier filter applied in QGIS to the original 795-point and 334-point datasets, respectively (Figure 5).

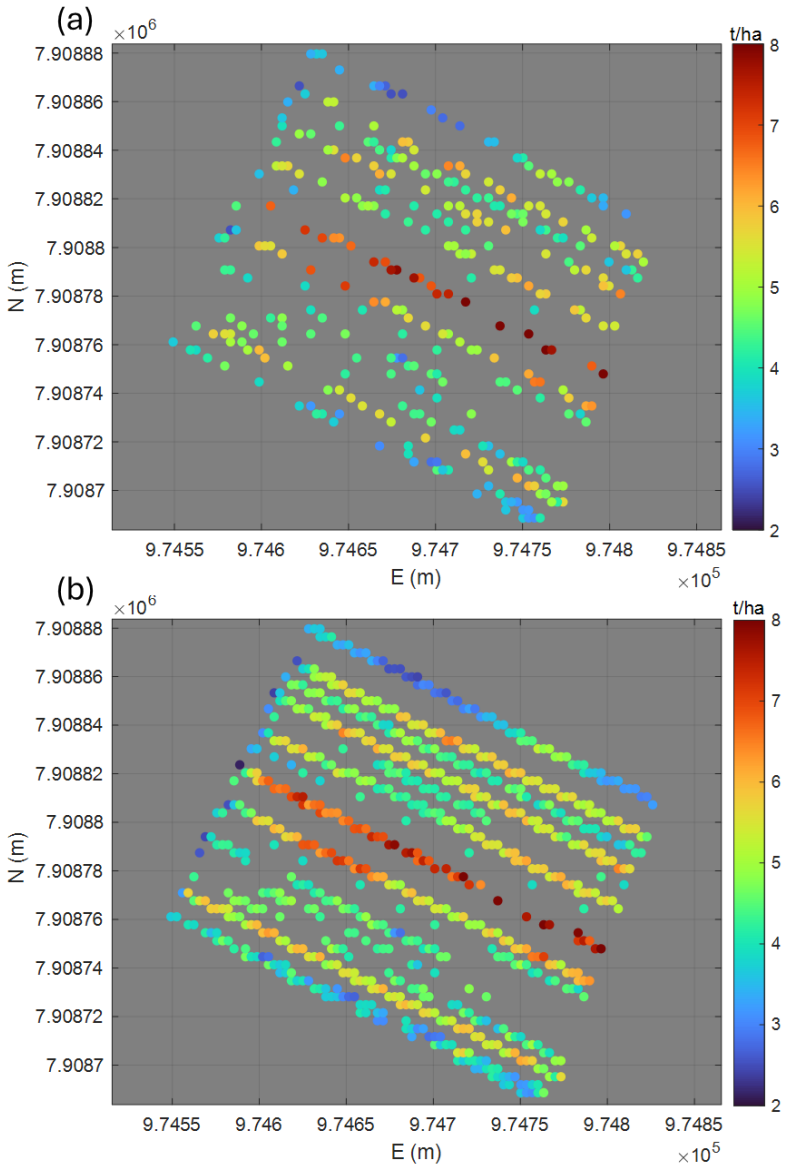

Figure 5: Randomly selected samples for GPR, OK, and SVM comparison: (a) 768 points, (b) 322 points.

The analysis relies on statistical metrics such as the coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to assess the

prediction accuracy of each method under different sampling scenarios. While RMSE and R^2 evaluate accuracy (including relative RMSE – rRMSE), MAE was also computed to capture the presence of systematic effects in the estimates. MAE is always lower than RMSE and it serves as an indicator of systematic biases, whereas RMSE accounts for both systematic and random errors.

RESULTS AND DISCUSSION

Evaluation of GPR Performance Under Different Sample Sizes

The results indicate that GPR maintained high performance in terms of accuracy for sample sizes down to 60% of the internal validation dataset, with $R^2 = 0.79$ and RMSE of 0.77 t/ha (rRMSE = 11.17%) for external validation. Additionally, MAE remained stable at approximately 0.35 t/ha for these scenarios (Figures 6, 7, 8, 9).

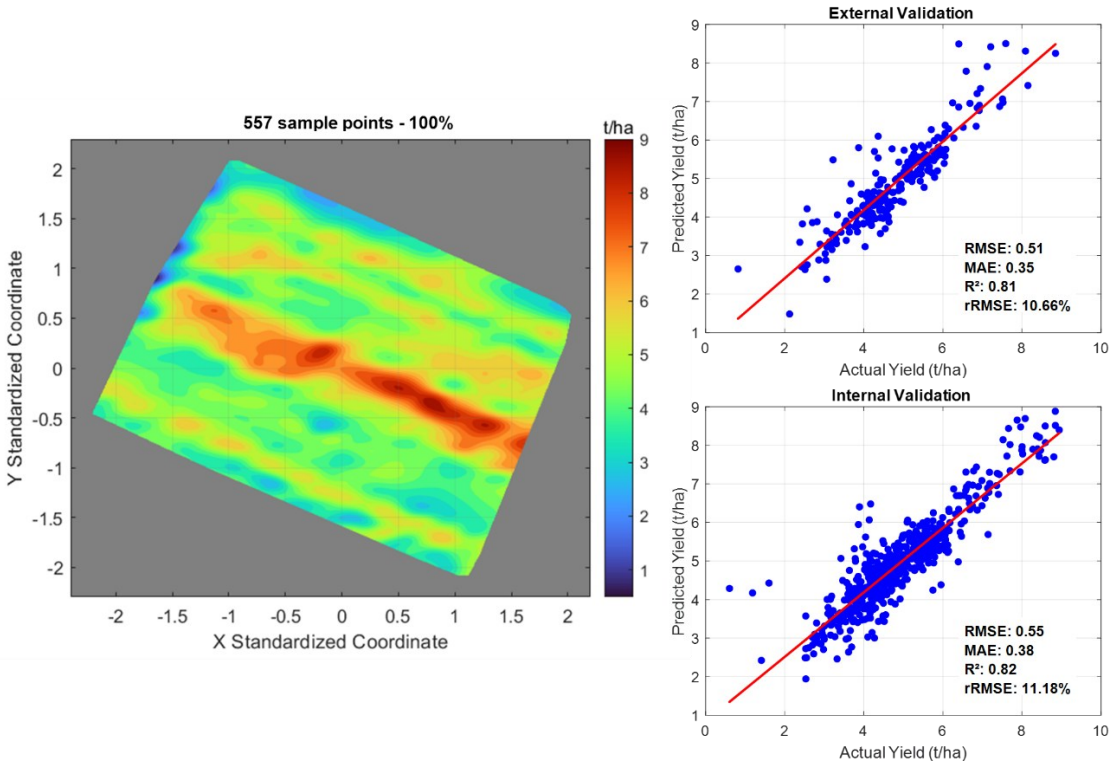

Figure 6: GPR-based yield spatial prediction and model performance using 557 samples.

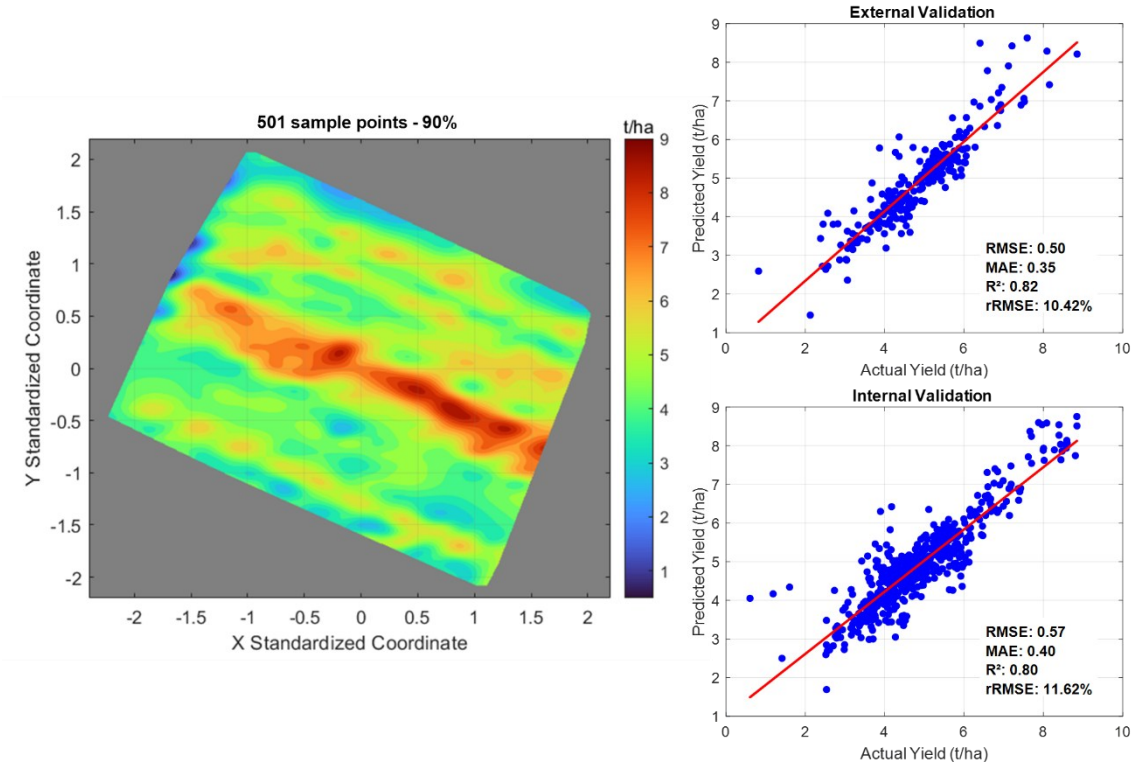

Figure 7: GPR-based yield spatial prediction and model performance using 501 samples.

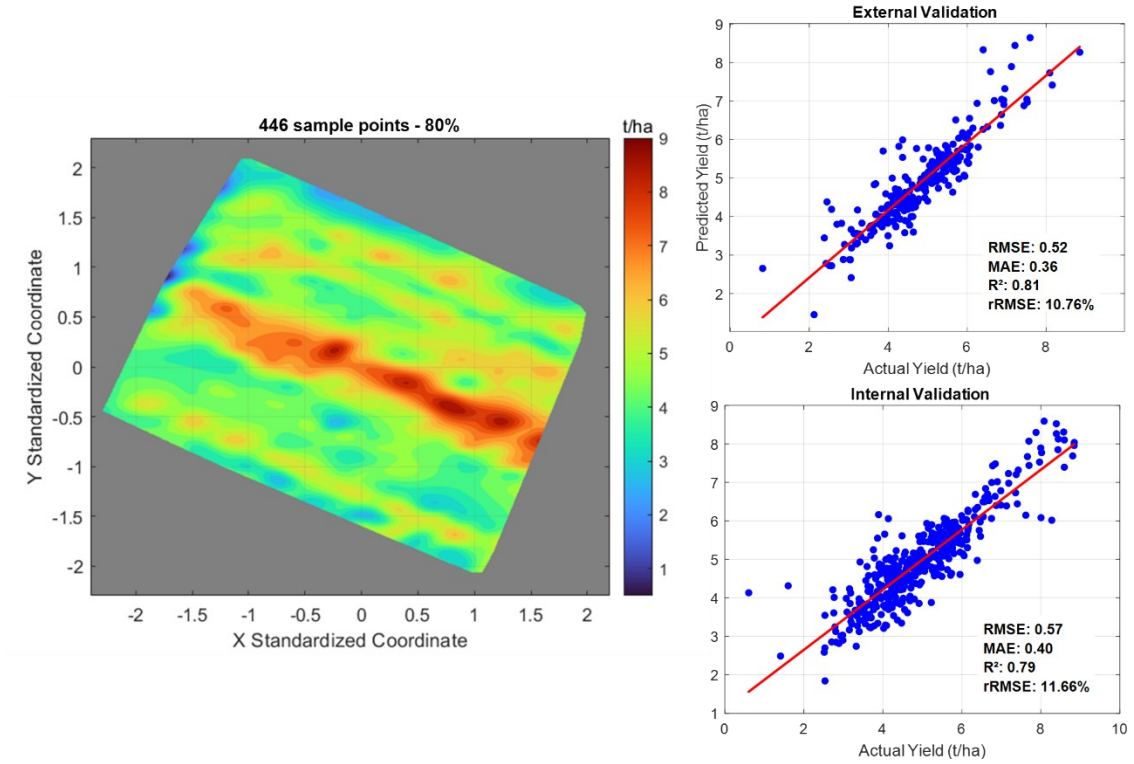

Figure 8: GPR-based yield spatial prediction and model performance using 501 samples.

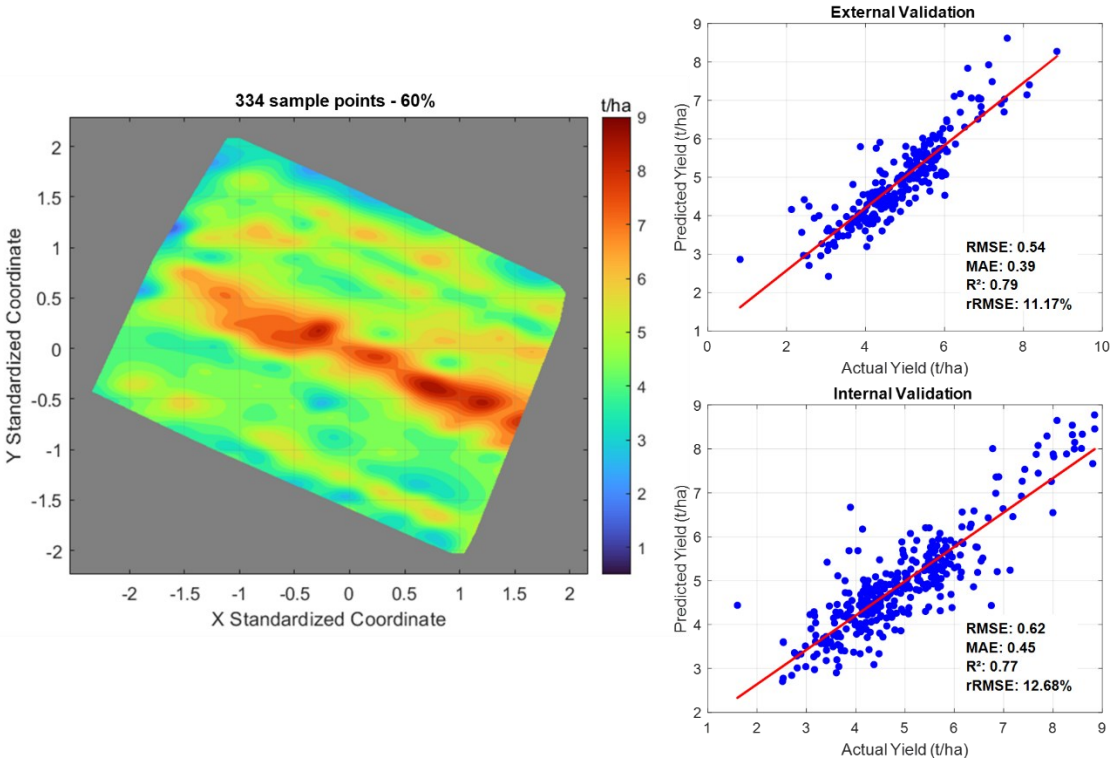

Figure 9: GPR-based yield spatial prediction and model performance using 334 samples.

For 40% of the internal validation set (223 points), GPR still performed reasonably well, with $R^2 = 0.69$, RMSE at 0.65 t/ha (rRMSE = 13.59%), and MAE at 0.48 t/ha (Figure 10). These results suggest that, although accuracy metrics indicate a slight decline with lower sample densities, GPR retains its ability to capture yield spatial patterns. This is consistent with findings from He et al. (2024) [9], who reported that GPR models provide excellent generalization ability even with small sample sizes.

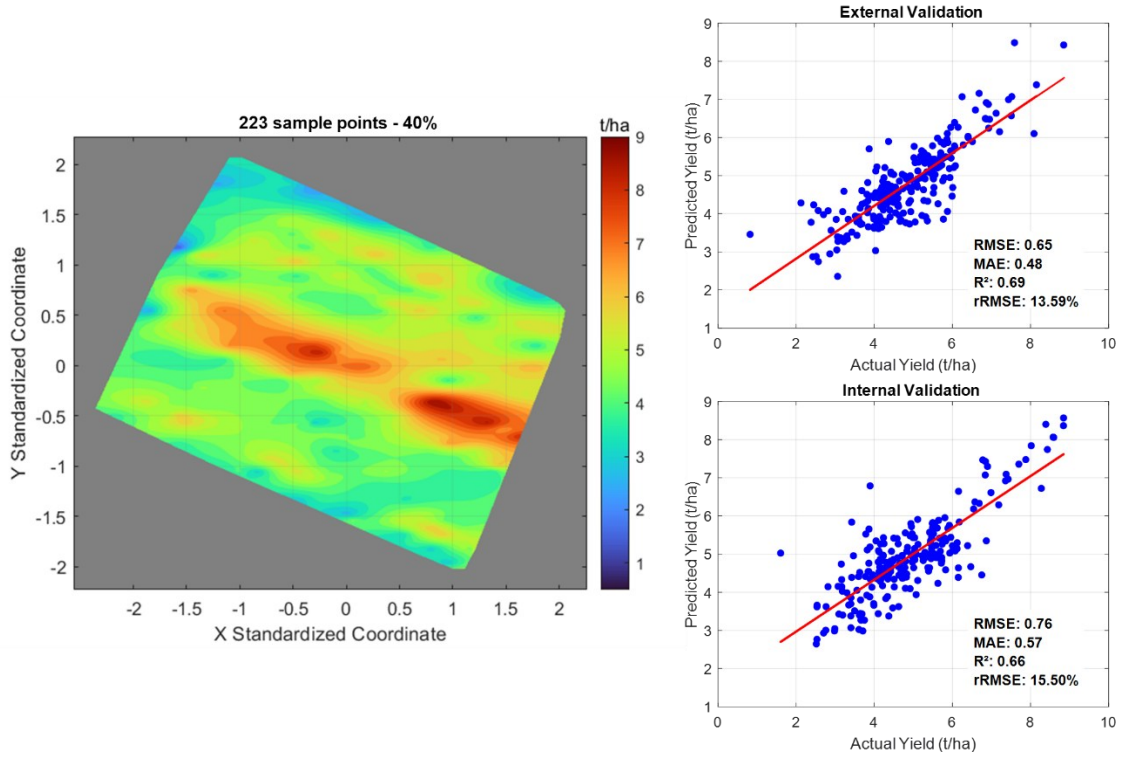

Figure 10: GPR-based yield spatial prediction and model performance using 223 samples.

The effect of data scarcity became more pronounced when only 112 points (20% of the internal validation dataset) were used. This limitation is particularly evident in localized patterns, where GPR's capacity to model finer spatial details diminishes (Figure 11). However, in this scenario, the main limitation was not the GPR itself, but the insufficient number of samples, which prevented the model from effectively capturing local yield variations. The reduced sample density led to an oversmoothing effect, where finer spatial details were lost. These results highlight the importance of optimizing sample sizes to balance the trade-off between capturing detailed local patterns (which requires a higher number of samples) and reducing operational costs (which benefits from lower field sampling efforts). This balance is crucial in precision agriculture applications, where economic constraints often limit extensive data collection, yet spatial heterogeneity must be accurately represented for effective decision-making.

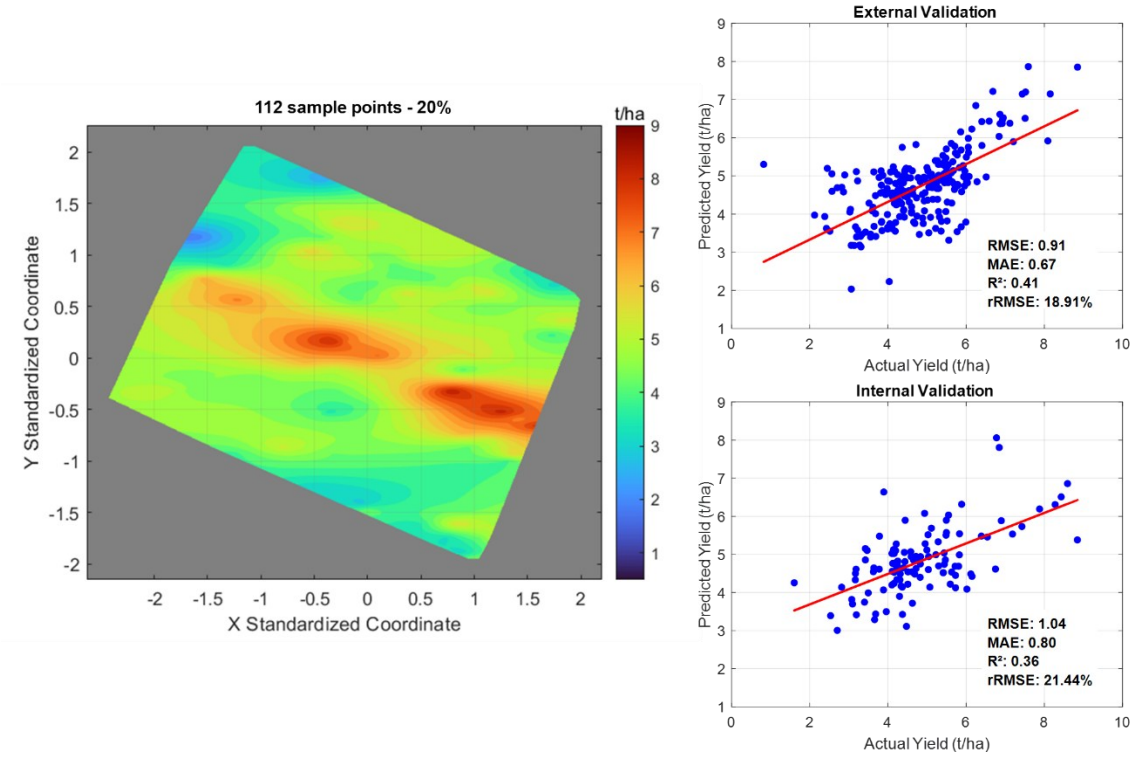

Figure 11: GPR-based yield spatial prediction and model performance using 112 samples.

Performance Evaluation of GPR, Ordinary Kriging, and SVM in Yield Prediction

Ordinary Kriging (OK) is the default interpolation method in the Smart-Map module. This method relies on the calculation of a semivariogram, which enables the analysis of the spatial dependence of soybean yield. To optimize performance, different semivariogram models were tested, and those with the lowest RMSE were selected. The Exponential model was chosen for both 768-sample dataset and 369-sample dataset (Figure 12). For the sample of 768 points, OK was performed with a search radius of 49.132 meters, considering up to 16 neighboring points for interpolation. The variogram model presented a nugget effect of 0.02, a sill of 1.268, and a coefficient of determination of 0.931, indicating a good model fit. The RMSE was 0.06, suggesting a low discrepancy between predicted and observed values. For the sample of 322 points, the results were like those obtained for 768 points, using the same exponential model but with different parameter values. In this case, the nugget effect was 0.06, the sill 1.455, and the range

59.144 meters. The coefficient of determination (0.945) remained high and RMSE low (0.063), indicating a low discrepancy between predicted and observed values.

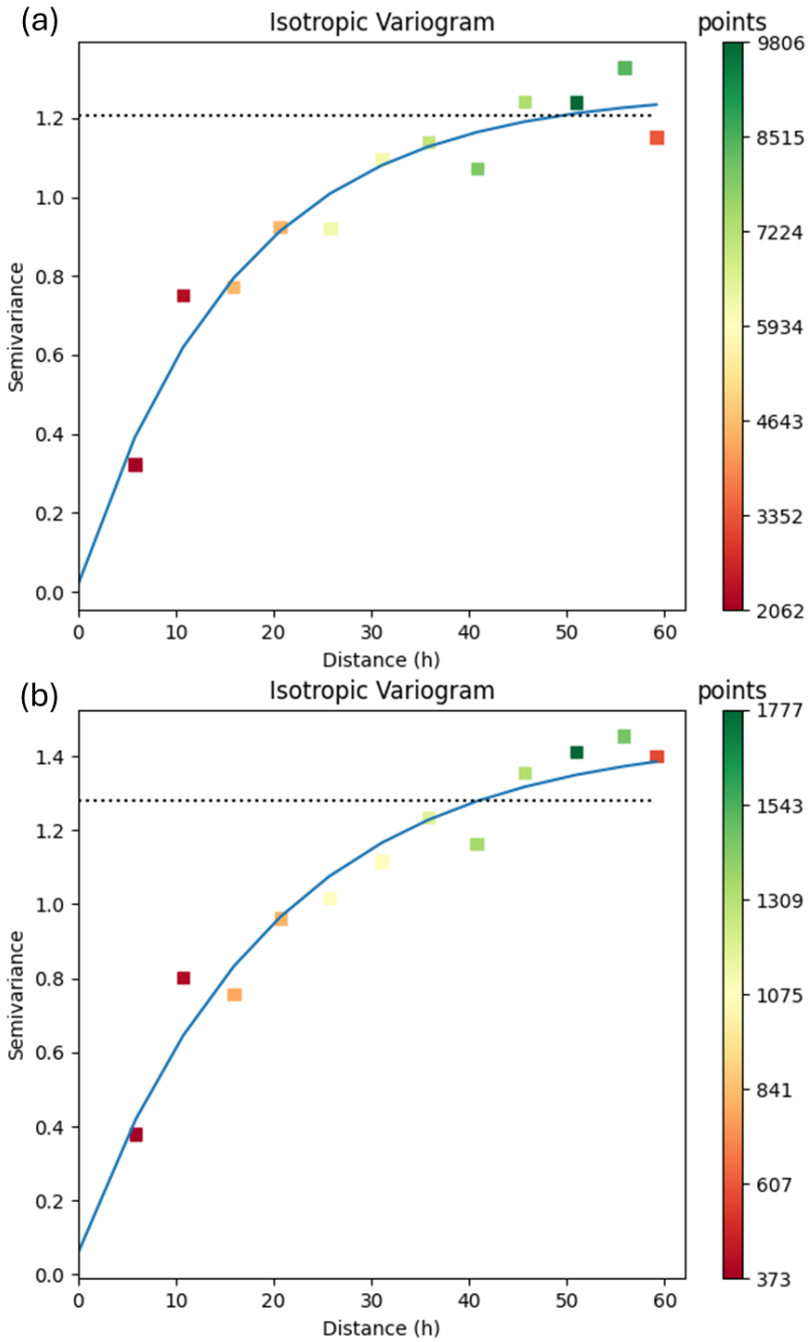

Figure 12: Semivariogram and fitted models: Exponential for both 768 sample points (a) and for 322 sample points (b).

For SVM, the choice of kernel function is a crucial hyperparameter. In the Smart-Map plugin, the Radial Basis Function (RBF) kernel is the default, as it has been found to be well-suited for most datasets due to its ability to capture nonlinear relationships [10].

The spatial distribution of soybean yield is influenced not only by different interpolation methods but also by the sampling density used to generate yield maps. The results displayed in Figure 13 indicated that there were no significant differences between Gaussian Process Regression (GPR), Ordinary Kriging (OK), and Support Vector Machine (SVM) in cross-validation, based on the analysis of regression plots and quality indicators (RMSE, MAE, R^2, and rRMSE). However, a smoothing effect was observed in the yield maps generated by SVM in Smart-Map.

The performance of GPR and OK remained similar for larger sample sizes, but GPR exhibited a slight advantage, particularly in scenarios with reduced data availability (Figure 14). As the number of sample points decreased, a smoothing effect became more apparent in the yield maps generated by SVM, suggesting lower sensitivity to local variations and a potential loss of critical spatial trends. This behavior can be attributed to the limited number of support vectors available for training, which prevents SVM from capturing finer-scale yield variations. When support points are too widely spaced, the resulting surfaces tend to be overly smoothed, leading to less detailed and less precise estimates [14]. Despite these limitations, SVM has evolved into an efficient paradigm for handling nonlinear problems [15], yet it struggles in spatial interpolation where spatial continuity is key.

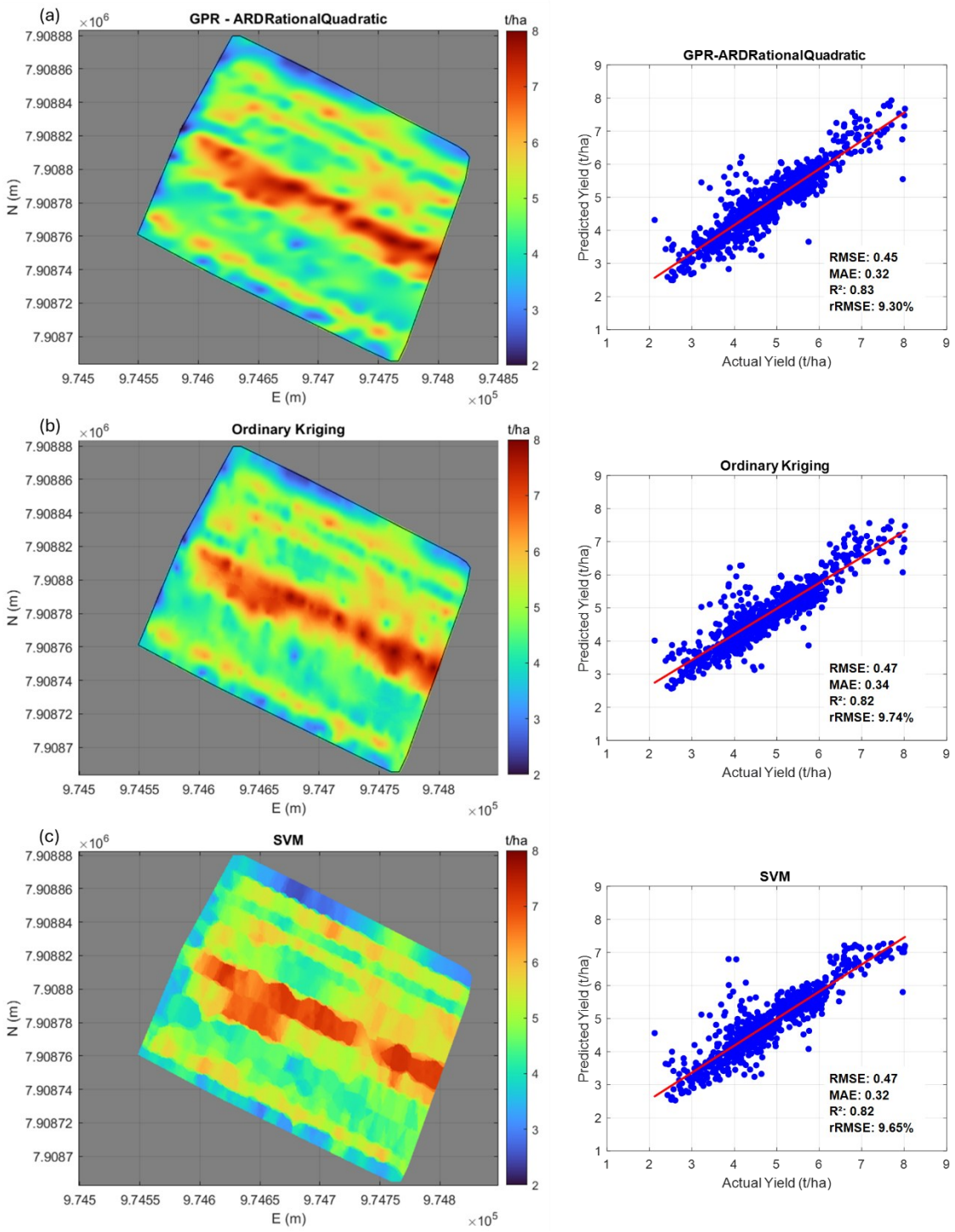

Figure 13: Predictive Performance of GPR (a), Ordinary Kriging (b), and SVM-RBF (c) for Soybean Yield Prediction at 768-sample points dataset.

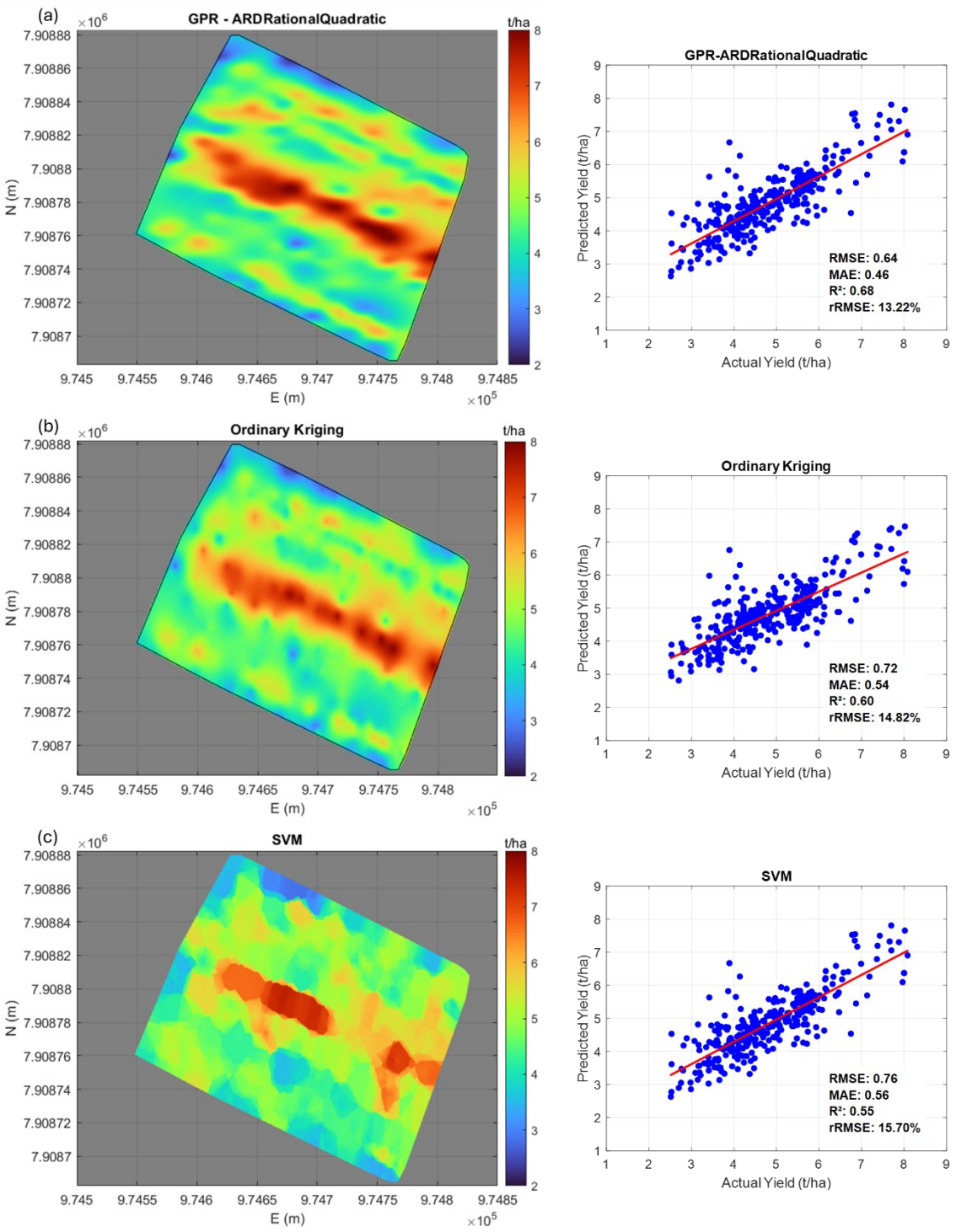

Figure 14: Predictive Performance of GPR (a), Ordinary Kriging (b), and SVM-RBF (c) for Soybean Yield Prediction at 322-sample points dataset.

OK leverages the spatial dependence structure of yield data through the semivariogram, making it a suitable method in sparse data conditions. However, GPR's structure, particularly with the Rational Quadratic Kernel with a separate length scale per predictor, ensures a higher level of detail in regions with fewer observations. GPR model fitting involves estimating covariance function parameters and noise variance, optimizing hyperparameters using maximum likelihood estimation (MLE). The model's kernel function plays a fundamental role in adjusting to different spatial patterns, allowing GPR to capture finer spatial details more effectively than OK, especially when the sampling density is reduced.

In scenarios where yield spatial distribution is highly variable due to biotic and abiotic factors, GPR emerges as a more robust alternative. For large sample sizes (e.g., 768 samples in this study area), GPR and OK performed similarly, both outperforming SVM. However, previous studies suggest that SVM may surpass OK in certain conditions, such as when dealing with high-dimensional feature spaces rather than pure spatial interpolation tasks [11, 16]. On the other hand, research has also shown that GPR offers noticeable gains in terms of accuracy and robustness when compared to other machine learning approaches for soybean yield estimation [5].

The differences observed between GPR, OK, and SVM in estimating soybean yield can be explained by various biotic and abiotic factors influencing spatial yield distribution in the field. Specifically, in this study area, soil variability and topography played crucial roles. The heterogeneity of soil properties, including organic matter content, water retention capacity, and nutrient availability in the Red Latosol, created complex spatial patterns that GPR captured more accurately due to its ability to model spatial

autocorrelation through kernel functions and optimized hyperparameters. Topography also influenced the distribution of soil moisture and nutrients, which was better represented by GPR and OK.

Additionally, microclimatic variations within the study area, such as differences in solar radiation and temperature, likely contributed to spatial yield patterns. GPR and OK effectively modeled these variations, whereas SVM tended to produce overly smoothed predictions, failing to capture localized fluctuations. Among the biotic factors that impacted yield estimates, pest and disease distribution, uneven crop growth, and past agricultural management practices played key roles [17]. The irregular presence of soybean pests, such as caterpillars, created patches of low productivity, which GPR and OK identified more effectively due to their reliance on spatial dependence modeling. Uneven crop growth, resulting from emergence variability and intraspecific competition, also introduced complex patterns that GPR and OK captured with greater accuracy.

Furthermore, historical agricultural practices, such as variable-rate fertilization and crop rotation, left spatial imprints on yield distribution, which GPR and OK detected more effectively, whereas SVM may require a larger dataset to learn these spatial relationships. The results suggest that GPR exhibited the least sensitivity to sample density reduction, followed by OK, while SVM suffered the most from data sparsity. This highlights the importance of selecting interpolation models that maintain stability under different sampling conditions, reinforcing that GPR is the most resilient option in this study. In soybean cultivation, characterized by its specific nutritional requirements and strong dependence on climatic conditions, the ability to accurately detect yield variations can mean the difference between a profitable harvest and significant losses [18].

From a practical standpoint, these findings confirm that GPR is a valuable tool for precision agriculture applications, particularly when sample availability is limited. The balance between interpolation accuracy and computational efficiency makes GPR a promising alternative for large-scale yield mapping. Furthermore, its kernel-based structure allows for flexible adaptation to different spatial patterns, ensuring robust predictions even in regions with fewer observations.

CONCLUSION

This study evaluates the impact of interpolation methods and sample size reduction on soybean yield prediction. GPR and OK consistently outperform SVM, with GPR showing the highest robustness to sample size reduction. OK remains effective but is more sensitive to sample density and relies on the assumption of isotropy, which may not always hold in agricultural fields. SVM in Smart-Map suffers the most from data sparsity, producing smoother estimates and failing to capture finer spatial variations. However, optimizing SVM with a more suitable kernel function could enhance its performance. From a practical perspective, GPR and OK remain strong candidates for yield interpolation, reinforcing the importance of selecting methods based on data availability and spatial variability to ensure accurate predictions in precision agriculture.

ACKNOWLEDGMENTS, FINANCIAL SUPPORT, AND FULL DISCLOSURE

The authors acknowledge the institutional support and resources provided for the development of this study. This work was funded by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) through a scholarship awarded to Samuel Phillipe. Additionally, this study received financial support from the National Council for Scientific and Technological Development (CNPq), process no. 421278/2023-4, linked to the second author.

The authors also express their gratitude to the institutions and collaborators who contributed with data, infrastructure, and technical support for the research.

Conflict of Interest Statement: The authors declare no conflict of interest in conducting and publishing this work.

REFERÊNCIAS

[1] Lucas MP, Longman RJ, Giambelluca TW, Frazier AG, McLean J, Cleveland SB, et al. Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain. J Hydrometeorol. 2022;23(4):561-572. doi:10.1175/JHM-D-21-0171.1.

[2] Sifaou H, Belmeguenai Y, Messaoudi D, Kannouf N, Chikhi S. A precise performance analysis of support vector regression. In: Proceedings of the 38th International Conference on Machine Learning (ICML); 2021 Jul 18-24; Virtual Conference. PMLR; 2021. p. 9749-59. Available from: http://proceedings.mlr.press/v139/sifaou21a/sifaou21a.pdf

[3] Siqueira, D. S., Barros, J. G. A., Marques, J. S., & Pereira, G. T. (2021). Geostatistical and machine learning models for spatial prediction of soil properties in precision agriculture. Revista Ceres, 68(5), 384-393. DOI: 10.1590/0034-737X202168050008

[4] Campos-Taberner M, García-Haro FJ, Moreno Á, Gilabert MA, Sánchez-Ruiz S, Martínez B, Camps-Valls G, et al. Mapping Leaf Area Index with a Smartphone and Gaussian Processes. IEEE Geosci Remote Sens Lett. 2015;12(12):2501-5. doi: 10.1109/LGRS.2015.2488682.

[5] Martínez-Ferrer L, Piles M, Camps-Valls G. Crop Yield Estimation and Interpretability with Gaussian Processes. IEEE Geosci Remote Sens Lett. 2021;18(12):2043-7. doi: 10.1109/LGRS.2020.3016140.

[6] Alebele Y, Wang W, Yu W, Cheng T, et al. Estimation of Crop Yield From Combined Optical and SAR Imagery Using Gaussian Kernel Regression. IEEE J Sel Top Appl Earth Obs Remote Sens. 2021;14:1-1. doi: 10.1109/JSTARS.2021.3118707.

[7] Martins GD, Xavier LCM, Oliveira GP, Gallo MLBT, Abreu Júnior CAM, Vieira BS, et al. Using geospatial information to map yield gain from the use of Azospirillum brasilense in furrow. Agronomy. 2023;13(3):808. doi:10.3390/agronomy13030808.

[8] Climate-Data.org. Climate of Patos de Minas. [Internet]. [S.l.]: Climate-Data.org; 2024 [cited 2024 Dec]. Available from: https://en.climate-data.org/south-america/brazil/minas-gerais/patos-de-minas-2893/

[9] He H, Zheng T, Zhao J, Yuan X, Sun E, Li H, et al. Improved Gaussian regression model for retrieving ground methane levels by considering vertical profile features. Front Earth Sci. 2024;12:1352498. doi: 10.3389/feart.2024.1352498

[10] Rasmussen, C. E. and C. K. I. Williams. *Gaussian Processes for Machine Learning.* MIT Press. Cambridge, Massachusetts, 2006.

[11] Pereira GW, et al. Smart-Map: An Open-Source QGIS Plugin for Digital Mapping Using Machine Learning Techniques and Ordinary Kriging. Agronomy. 2022;12(6):1350. DOI: 10.3390/agronomy12061350.

[12] Rodrigues BP, Rofatto VF, Matsuoka MT, Teles Assunção T. Resampling in neural networks with application to spatial analysis. Geo-Spat Inf Sci. 2022;25(3):413–24. doi:10.1080/10095020.2022.2040923.

[13] Li J. A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R. Appl Sci. 2019;9(10):2048. doi:10.3390/app9102048.

[14] Vapnik VN. The nature of statistical learning theory. 2nd ed. New York: Springer; 2000.

[15] Chandra MA, Bedi SS. Survey on SVM and their application in image classification. Int J Inf Technol. 2018. doi: 10.1007/s41870-017-0080-1.

[16] Ribeiro JL, Silva AM. Comparative study of kriging and support vector machines for spatial prediction of soil properties. Computers and Electronics in Agriculture. 2018;155:183-193.

[17] Carvalho E, Assis GA, Martins GD, Marques DJ, Santos EA, Xavier LCM, et al. Intercropped soybean plant population in a coffee plantation and its effects on agronomic parameters and geospatial information. Agronomy. 2024;14(2):343. doi:10.3390/agronomy14020343.

[18] Filippi P, Han SY, Bishop TFA. On crop yield modelling, predicting, forecasting and addressing the common issues in published studies. Precis Agric. 2024;26(8). doi:10.1007/s11119-024-10212-2.

AUTHOR CONTRIBUTIONS

Conceptualization: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530

Data curation: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325

Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

Formal analysis: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Samuel Phillippe https://orcid.org/0009-0001-3020-4595

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Investigation: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

Methodology: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

George Deroco Martins https://orcid.org/0000-0001-9738-7325

Software: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530.

Resources: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530

George Deroco Martins https://orcid.org/0000-0001-9738-7325.

Supervision: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325.

Validation: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

Visualization: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

Writing – original draft: Samuel Phillippe https://orcid.org/0009-0001-3020-4595.

Writing – review & editing: Vinicius Francisco Rofatto https://orcid.org/0000-0003-1453-7530;

George Deroco Martins https://orcid.org/0000-0001-9738-7325;

This preprint was submitted under the following conditions:

- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a Creative Commons CC-BY license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.