Machine Learning Algorithms for Spatial Interpolation of Crop Yield at the Field Scale

Franco Suarez

suarezfranco@agro.unc.edu.ar

Unidad de Fitopatología y Modelización Agrícola (UFyMA), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Córdoba. Argentina. Ing. Agr. Félix Aldo

Pablo Paccioretti

Unidad de Fitopatología y Modelización Agrícola (UFyMA), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Córdoba. Argentina. Ing. Agr. Félix Aldo

Mónica Balzarini

Unidad de Fitopatología y Modelización Agrícola (UFyMA), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Córdoba. Argentina. Ing. Agr. Félix Aldo

Mariano Córdoba

Unidad de Fitopatología y Modelización Agrícola (UFyMA), Instituto Nacional de Tecnología Agropecuaria (INTA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Córdoba. Argentina. Ing. Agr. Félix Aldo

---

Research Article

Keywords: Model performance evaluation, Geostatistical methods, Spatial variability, Predictive modelling, Neural networks

Posted Date: May 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9727868/v1

Additional Declarations: No competing interests reported.

MACHINE LEARNING ALGORITHMS FOR SPATIAL INTERPOLATION OF CROP YIELD AT THE FIELD SCALE

Franco Suarez[a]*, Pablo Paccioretti[a], Mónica Balzarini[a] and Mariano Córdoba[a]

[a] Grupo de vinculado de Estadística y Biometría, Unidad de Fitopatología y Modelización Agrícola (UFyMA) Instituto Nacional de Tecnología Agropecuaria (INTA). Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Argentina. Ing. Agr. Félix Aldo Marrone 746 (X5000HRA), Córdoba, Argentina.

*Corresponding autor: suarezfranco@agro.unc.edu.ar

HIGHLIGHTS

- Coupled with spatial covariables, machine learning showed low prediction error in spatial interpolation of yield monitor data
- Machine learning methods are faster than kriging in large datasets.
- Quantile regression forest presented lower prediction error and provided direct measures of prediction uncertainty

IMPACT

Through the systematic evaluation of spatial machine learning algorithms on over 1,000 real-world yield monitor datasets, we identified Quantile Regression Forest (QRF) as a robust and computationally efficient tool for in-field yield mapping, improving both the accuracy of spatial prediction and the quantification of uncertainty. These results provide practical guidance to agronomists and data scientists seeking to improve resource use efficiency and site-specific management in crop production systems. This study directly supports the precision agriculture

goal of collecting, processing, and analyzing spatial data to improve management decisions based on estimated variability.

KEYWORDS

Model performance evaluation, Geostatistical methods, Spatial variability, Predictive modelling, Neural networks

ABSTRACT

Crop mapping is one of the most frequent uses of yield monitor data in precision agriculture. Processing such data requires spatial interpolation to predict yields at unsampled locations. Ordinary kriging (OK) is the geostatistical interpolation method usually employed for mapping georeferenced data. In recent years, machine learning (ML) algorithms have gained attention for spatial interpolation tasks because of their ability to process large data volumes; however, their comparative performance for yield mapping at the field scale remains limited. This study compares the statistical behavior of ML methods, including quantile regression forest (QRF), generalized boosted regression models (GBM), extreme gradient boosting (XGB), and radial basis function neural networks (RBFNs) for processing yield monitor data. OK is applied as reference. To enhance the performance of ML algorithms for fine-scale yield mapping, covariates from the spatial neighborhood of sampled yield values were incorporated to predict yields at unsampled sites. Over 1,000 yield monitor datasets from multiple crop species were processed to assess algorithm performance. The ML methods—QRF, GBM, and XGB—demonstrated robust statistical performance, since they effectively handled large data volumes and improved spatial interpolation accuracy. The QRF method achieved the highest error reduction (on average, an 8.7% error reduction), was faster than OK, and generated high-quality uncertainty maps. GBM and XGB also

performed better than OK. Coupled with spatial covariables, the studied ML algorithms are valuable alternatives to conventional kriging for yield mapping at the field scale.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 INTRODUCTION

Crop yield mapping, utilizing data from yield monitors attached to automated harvesters with satellite control systems, is a widely adopted practice in modern agriculture (Bullock et al., 2019). These monitors enable the collection of yield data from numerous sites within a field, resulting in datasets with several observations per hectare. Processing such data entails data filtering, followed by spatial interpolation to predict yields at unsampled locations, and subsequently generating nearly continuous spatial yield maps (Cordoba & Balzarini, 2021). Yield maps are crucial for implementing site-specific management practices in precision agriculture (PA). Spatial prediction is widely used to explore patterns in the spatial distribution of phenomena (Bronowicka-Mielniczuk & Mielniczuk, 2023). Accurate predictions enable the analysis of spatial variability, determination of patterns, and characterization of phenomena across geographical spaces (Siabato et al., 2019). Geostatistics interpolators estimate the value of an unmeasured variable at a specific location using surrounding observed data while accounting for spatial autocorrelation. Spatial autocorrelation quantifies the relationship between variable values at different geographical sites within the field. Spatial autocorrelation is assessed by calculating the linear correlation between a variable at a specific site and its values at other neighboring sites, accounting for spatial lags. Positive spatial autocorrelation occurs when nearby observations show similar values, suggesting non-random spatial structures. Geostatistical models, such as Ordinary Kriging (OK) (Kerry et al.,

2010), are frequently employed for spatial predictions to characterize within-field spatial variability (Salem et al., 2024; Silva et al., 2023). These approaches require a statistical model to explain spatial variation, and explicitly incorporate spatial autocorrelation, which is crucial to model spatially structured stochastic process. Geostatistical interpolators also allow us to quantify prediction uncertainty, which is important for decision-making processes requiring uncertainty maps (Wadoux et al., 2020). However, geostatistical models could fail to explain non-stationary yield variability, since yield monitor data carries significant uncertainty due to uncontrolled factors such as measurement errors (Taylor et al., 2010), genotype-environment interactions, and soil quality, all of which contribute to unexpected yield variability (Koutsos et al., 2023). OK is an unbiased linear interpolation technique designed for stationary phenomena with unknown means. It incorporates a semivariogram model to represent spatial autocorrelation measuring both the scale and magnitude of spatial variability, providing parameters for interpolation. Proper semivariogram fitting is crucial, since parameter estimations significantly influence predictions (Giraldo et al., 2023; Oliver et al., 2014). OK is computationally intensive for large datasets and relies on strict statistical assumptions, such as data normality and stationarity (Cordoba & Balzarini, 2021).

In recent years, machine learning (ML) algorithms have emerged as promising tools for spatial prediction, particularly in digital soil mapping (e.g.,Wadoux et al., 2020) and PA using statistical covariates (e.g., Burdett & Wellen, 2022). ML methods refer to a class of data-driven, non-linear algorithms primarily used for information extraction and pattern recognition. One key advantage of ML algorithms is their ability to model complex interactions among data, and for many of them, without assuming any predefined distribution (Wadoux et al., 2020). As computational capabilities increased, ML has become a promising alternative across data analysis in agriculture. Although

ML such as Random Forest (RF) (Breiman, 2001), has proven effective, algorithms rarely incorporate spatial autocorrelation in the analysis. RF and several other tree-based ML algorithms, such as Generalized Boosted Regression Model (GBM) (Ridgeway, 1999), and Extreme Gradient Boosting (XGB) (Chen & Guestrin, 2016), have shown high predictive performance in regional-scale mapping (Wadoux et al., 2020). However, when making predictions these ML methods ignore spatial information from neighboring sites.

Among ML methods, neural networks (NN) have become widely used for various tasks such as interpolation, function approximation, classification, and time series prediction (Imanian et al., 2023). Radial Basis Function Networks (RBFNs) are a type of neural network applied in various domains, including interpolation and prediction (Imanian et al., 2023). Despite their widespread use, spatially constrained versions of neural networks have not been used for yield mapping. Although several ML techniques are used in agriculture (Sergeev et al., 2019), few have been specifically tested for intra-field yield mapping, especially those that account for spatial autocorrelation.

To address this limitation, some spatially constrained ML models have been developed. One approach to capture spatial autocorrelation involves adding kriged residuals to the ML prediction at each location (Li et al., 2011). Other strategies include incorporating geographic covariates, such as longitude and latitude (Meyer et al., 2019), buffer distance maps from observation points (Hengl et al., 2018), or geographically weighted covariables (Georganos et al., 2019). Córdoba and Balzarini (2021) used a variant of RF, called Quantile Regression Forest (QRF) (Meinshausen et al., 2006), as a base method in the implementation of a spatial interpolation algorithm that incorporates the autocorrelation of georeferenced data. This method explicitly integrates geographical context into ML by including actual observations at nearby locations as covariates.

In this study, we compared the predictive performance of ML methods that explicitly incorporate geographical context, including QRF, GBM, XGB, and RBFNs, along with OK method. All ML methods included neighboring observations and their distances as covariates, to assess spatial variability of crop yield at the field scale.

2 MATERIALS AND METHODS

2.1 Data

Spatially adjusted ML algorithms were run using 1050 yield maps; the datasets were collected from extensively cultivated agricultural fields located in the Argentine Pampas region, specifically in the provinces of Buenos Aires and La Pampa. Based on the harvested crop, the datasets from yield monitors were classified as follows: soybean (*Glycine max* (L.) Merr.) (48%), corn (*Zea mays* L.) (24%), wheat (*Triticum* spp.) (13%), barley (*Hordeum vulgare*) (9%), and sunflower (*Helianthus annuus* L.) (7%). The average field size was 58 ha, and the mean number of observations per field or dataset was n = 9844. Yield monitor data were pre-processed according to the protocol proposed by Vega et al. (2019).

2.2 Incorporation of spatial information in ML algorithms

To incorporate spatial information into the ML algorithms, several covariates were calculated using observed yield data from the n nearest sites and their respective distances to the target sites for spatial prediction. The value of n was set to six observations. This process was performed using the near.obs function (Sekulić et al., 2020) from the meteo package in R software (R Core Team, 2024).

The predicted yield at a target location S_0, denoted as $\hat{Y}_{S_0}$, was obtained using the yield observations from the n nearest neighboring sites. For each neighbor S_i, we collected both the observed yield value $X_{(S_i)}$ and the corresponding distance d_i to the target location. The prediction model integrates these two sources of information—yield and distance—through a function f that maps them into the predicted yield:

$$\hat{Y}_{S_0} = f(X_{(S_1)}, d_1, X_{(S_2)}, d_2, \ldots, X_{(S_n)}, d_n)$$

where $X_{(S_i)}$ denotes the observed yield at neighboring site S_i, and d_i represents its distance from the target site S_0. The function f specifies the prediction model, which combines spatial distance with observed yield information to estimate yield at the target location.

2.3 Parametrization of compared algorithms

2.3.1 Ordinary Kriging (OK)

The OK method was implemented with the aim of obtaining a reference method. We used the local OK variant (kriging neighborhood), which is recommended for large datasets such as those derived from yield maps (Oliver & Webster, 2014; Schabenberger & Gotway, 2017). A key feature of local OK is the use of a limited number of neighboring points for predictions, ensuring that distant observations have minimal influence. This approach supports the assumption of local stationarity without considering large-scale fluctuations, thus enhancing the applicability of this geostatistical interpolation method (Karapetsas et al., 2022).

In this study, two configurations of the number of neighbors (6 and 25) were evaluated, referred to as OK_6 and OK_25. For OK_6, the *krige* function was used, setting the nmax parameter

(maximum number of nearest neighbors for prediction) to six. For OK_25, the same function was employed, but with nmax set to 25, following the guidelines of Oliver & Webster (2014). Both models were implemented using the gstat package (Pebesma & Graeler, 2012). The semivariogram model and its parameters were estimated using the *autofitVariogram* function from the automap package (Hiemstra & Hiemstra, 2013). The semivariogram models tested included spherical (Sph), exponential (Exp), Gaussian (Gau), and the M. Stein parameterization (Ste) (Minasny & McBratney, 2005). All analyses were conducted in R software (R Core Team, 2024).

2.3.2 Quantile Regression Forest (QRF)

QRF is an extension of the Random Forest (RF) algorithm, designed to overcome some limitations inherent in standard RF models when data sets have important outliers. While RF is primarily a point prediction model (Dang et al., 2022), providing an estimate of the conditional mean of the output variable by averaging the predictions of numerous decision trees, it does not directly account for the uncertainty in those predictions. In contrast, QRF modifies the RF framework to retain a fuller range of information. Instead of only outputting a single point estimate (the mean prediction), QRF keeps track of all individual predictions made at each leaf node across the forest. This enhancement allows QRF to estimate the entire conditional distribution of the predicted variable, not just a central tendency measure. This enables the generation of both spatial predictions and uncertainty maps (Kirkwood et al., 2016).

The QRF method was implemented using the ranger package (Wright et al., 2017) of R software (R Core Team, 2024). The mtry hyperparameter, which specifies the number of covariates used to construct each regression tree, was set to one-third of the total covariates. Additionally, the ntree hyperparameter, which defines the number of trees to train, was set to 200. For QRF, the quantreg function argument was enabled by setting it to TRUE, which activated the

quantile regression functionality, allowing for the estimation of quantiles in the prediction distribution

2.3.3 Generalized Boosted Regression Model (GBM)

GBM captures nonlinear relationships between the response variable and a set of covariates by minimizing a specific loss function. It combines two existing techniques: decision trees and gradient boosting. The model focuses on the hard-to-learn samples by adding additional trees as iterations progress. This approach helps reduce the loss function, which measures the predictive performance of the model and its ability to predict the desired outcome. Specifically, the loss function penalizes large prediction errors more strongly, while assigning less weight to small errors (Al-Mudhafar et al., 2022).

GBM was tuned using the gbm package (Greenwell et al., 2019) in R software (R Core Team, 2024). The following hyperparameters were set: *distribution* was set to "gaussian", *n.trees* (the number of trees to train) was set to 100, *interaction.depth* (which specifies the maximum depth of each tree, or the number of variable interactions allowed) was set to one, *shrinkage* (the learning rate) was set to 0.1, and *n.minobsinnode* (the minimum number of observations in the terminal nodes of each tree) was set to 10.

2.3.5 Extreme Gradient Boosting (XGB)

XGB uses the gradient boosting method to improve the accuracy of predictions by sequentially building new models. The process begins with an initial model, which makes predictions for all observations in the dataset. Then, a specific loss function, such as mean squared error (MSE), is computed to assess the accuracy of the predictions. Based on this loss function, a new model is

adjusted and added to the ensemble to further reduce the loss. This process continues until no further significant improvements in prediction accuracy can be made (Al-Mudhafar et al., 2022).

XGB was tuned using the train function from the caret package (Max Kuhn, 2021) in R software (R Core Team, 2024), with the method set to xgbTree, which fits the model using the xgboost package (Chen et al., 2019). The following hyperparameters were used: nrounds (the number of iterations to perform before stopping the tuning process), which was set to 100; max_depth (the number of decision tree splits used during training), which was set to six; and eta (the learning rate), which was set to 0.3. A higher value of eta reaches the minimum of the objective function more quickly, resulting in a better model, but it may overshoot the optimal value. Conversely, a smaller value of eta will never reach the optimal value of the objective function, even after many iterations. The gamma hyperparameter (the minimum reduction in loss required to make an additional partition at a leaf node) was set to zero, and colsample_bytree (the fraction of columns sampled for constructing each tree) was set to one.

2.3.5 Radial Basis Function Networks (RBFNs)

RBFNs are a type of artificial neural network that use radial basis functions as activation functions in the hidden layer. Unlike traditional neural networks, which often rely on nonlinear activation functions like sigmoid or ReLU (Rectified Linear Unit), RBFNs use a radial basis function (typically Gaussian) to model the relationship between input variables and the target output. This structure allows RBFNs to capture complex nonlinearities in the data while offering advantages in terms of optimization speed and accuracy. The RBFN architecture involves three layers: an input layer, a hidden layer, and an output layer. The input layer consists of neurons corresponding to the number of features (i.e. input variables) in the dataset. The hidden layer, which contains the radial basis function activations, performs nonlinear transformation on the input

data, mapping it into a higher-dimensional space. The output layer produces a linear combination of these transformed features to make predictions. One of the key advantages of RBFNs is their relatively fast convergence during optimization compared to other types of neural networks, particularly when dealing with small to medium-sized datasets (Imanian et al., 2023). In this study, the RBFN model consisted of 12 input neurons, corresponding to six performance values from neighboring points and the six associated distances from these neighbors to the point being predicted. The output layer had a single neuron because the model was designed to predict a single performance value. The number of neurons in the hidden layer, which represent the centroids of the radial basis functions, is a critical hyperparameter that controls the complexity of the model. In this case, the number of neurons in the hidden layer was set to five, which is the default value for this function. The model was tuned using the train function from the caret package (Max Kuhn, 2021) in R software (R Core Team, 2024), with the *method* parameter set to "rbf". This allowed the model to be fitted using the RSNNS package (Bergmeir & Benítez Sánchez, 2012), which specializes in training RBFN models.

2.4 Statistical comparison criteria

The performance of the compared methods was evaluated using k-fold cross-validation (Hastie et al., 2009) with $k = 10$. In this procedure, the data were randomly partitioned into k groups; the model was then trained using $k-1$ groups, while the remaining group was used to assess the predictions. This validation was implemented for each of the 1050 datasets. The summary metrics for assessing predictive performance were: the root mean square error relative to the observed mean (nRMSE) (Karunasingha, 2022), the coefficient of determination (R^2) (Piepho, 2019), and Lin's concordance correlation coefficient (CCC) (Lin, 1992), which measures how well a dataset aligns with a reference measurement or test. A linear mixed model (Laird & Ware, 1982) was fitted

to compare nRMSE values, with interpolation methods and crop types as classification criteria and CV as a covariate. The model included interpolation methods, CV, and their interaction as fixed effects, while crop type was treated as a random effect.

To generate the probability distributions of prediction realizations for each model at each georeferenced point, a bootstrapping approach was employed (Gomes et al., 2019). These distributions enable the quantification of model uncertainty by calculating a prediction interval at a specific confidence level. In this approach, 80% of the data was used to fit the models, and the remaining 20% was reserved for validation. Bootstrapping involves repeated random sampling with replacement from the available data. A model is fitted to each bootstrap sample, which can then be applied to generate a map. By repeating the random sampling process and applying the model, probability distributions of prediction realizations for each model are produced.

To quantitatively evaluate uncertainty estimates, prediction interval coverage probability (PICP) plots were used. These plots assess whether the assigned probability to prediction intervals matches the empirical frequency of test data falling within those intervals (Schmidinger and Heuvelink, 2023). While the QRF and KG methods inherently allow for deriving prediction uncertainty, bootstrapping was applied to standardize uncertainty estimation across the compared methods.

3 RESULTS

The average yields from the yield maps were 8.95 t ha^{-1}, 3.97 t ha^{-1}, 3.79 t ha^{-1}, 2.97 t ha^{-1} and 2.63 t ha^{-1} for corn, barley, wheat, sunflower, and soybean, respectively. Table 1 summarizes the characteristics of the different crops in terms of average yield, field point density, and the CV, which reflects within-field variability. Regarding the CV, corn had the highest CV at 33%,

followed by barley (32%), soybean (31%), sunflower (28%), and wheat (26%). The minimum and maximum CV values observed across all yield maps ranged between 9% and 80%, indicating a wide range of yield variability within fields. Point density, a proxy for the spatial resolution of yield data, varied significantly across crops. Wheat had the highest variability in point density, ranging from 33 to 1002 points/ha, highlighting differences in sampling intensity across fields. Conversely, soybeans had a narrower range in point density, with a minimum of 53 and a maximum of 633 points/ha.

Table 1. Summary statistics of yield maps per crop. Values are means with minimum and maximum in parentheses.

Crop	Yield maps (#)	Field point Density (data/ha)	Yield (t ha^{-1})	Coefficient of Variation (%)
Barley	90	230 (91 – 924)	3.97 (0.6 – 7.3)	32 (16– 57)
Sunflower	73	175 (62 – 521)	2.97 (1.2 – 4.7)	28 (15 – 51)
Corn	250	227 (54 – 858)	8.95 (1.8 – 14.4)	33 (11 – 80)
Soybean	502	173 (53 – 633)	2.63 (0.5 – 5.7)	31 (10 – 80)
Wheat	135	226 (33 – 1003)	3.79 (1.1 – 6.3)	26 (9 – 52)

Figure 1 shows boxplots of the nRMSE distribution for each evaluated method. The nRMSE values for most methods were between 10% and 19%, except for RBFNs, which ranged between 16% and 25%. RBFNs exhibited the highest average nRMSE and the greatest variability in performance.

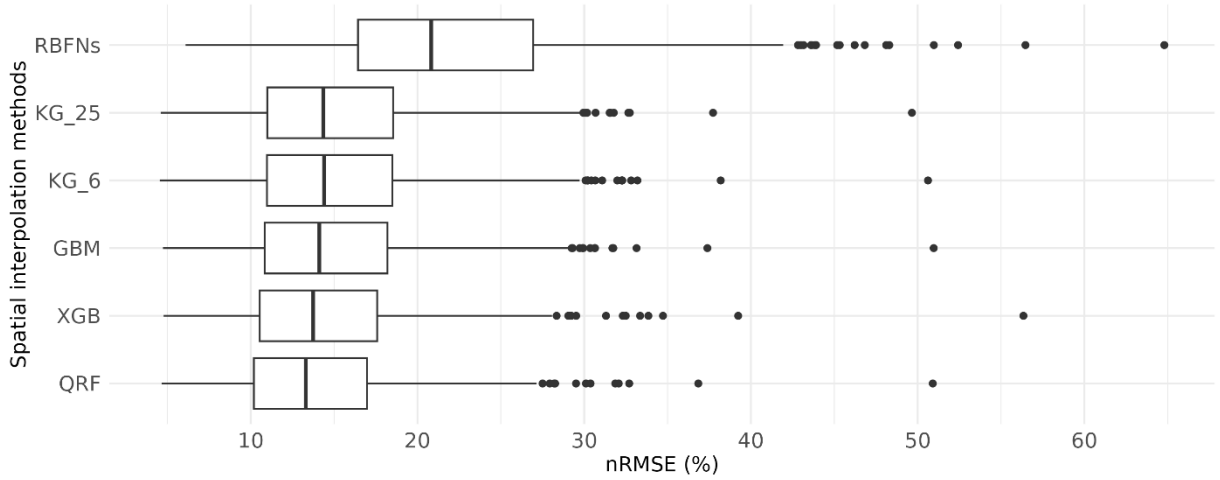

Figure 1. Boxplot of the nRMSE distribution for each evaluated method. GBM: generalized boosted regression model, KG_6: Ordinary Kriging using 6 neighboring observations, KG_25: Ordinary Kriging using 25 neighboring observations, RBFNs: radial basis function networks, QRF: quantile regression forest, XGB: extreme gradient boosting.

Table 2 shows the average nRMSE values for each method and the relative differences compared to OK_25, which served as the reference method. For all crops, QRF achieved the lowest nRMSE, followed by XGB, GBM, and RBFNs. Compared to OK_25, the average prediction error was reduced by 2% with GBM, 9% with QRF, and 6% with XGB, while it increased by 46% with RBFNs. For barley, the average nRMSE was 15%; for sunflower, it was 14%; soybean and maize had an average nRMSE of 16%, and wheat exhibited the lowest nRMSE at 12%. Regarding the number of maps where each method demonstrated the best predictive performance, QRF achieved the lowest nRMSE in 80% of the maps, followed by OK_25 (8%), OK_6 (5%), XGB (3%), GBM (2%). RBFNs did not perform best in any of the evaluated maps.

Table 2. Normalized root means square error relative to the observed mean (nRMSE, %) of six spatial interpolation methods. In parentheses: the relative difference values of each method compared to the ordinary kriging method using 25 neighboring observations for prediction (KG_25).

Crop	**nRMSE (Relative differences of nRMSE compared to KG_25)**

	GBM	RBFNs	QRF	XGB	KG_6	KG_25
Barley	14.2 (-3.4)	20.9 (42.2)	13.4 (-8.8)	13.9 (-5.4)	14.6 (-0.7)	14.7
Sunflower	13.6 (-2.2)	20.2 (45.3)	12.7 (-8.6)	13.2 (-5.0)	13.8 (-0.7)	13.9
Corn	15.7 (0.0)	23.6 (50.3)	14.8 (-5.7)	15.3 (-2.5)	15.7 (0.0)	15.7
Soybean	15.4 (-2.5)	22.9 (44.9)	14.5 (-8.2)	15.0 (-5.1)	15.9 (0.6)	15.8
Wheat	11.7 (-4.9)	17.3 (40.7)	10.8 (-12.2)	11.1 (-9.8)	12.1 (-1.6)	12.3
Average	14.1 (-2.6)	20.9 (44.6)	13.2 (-8.7)	13.7 (-5.5)	14.4 (-0.7)	14.5

GBM: generalized boosted regression model, RBFNs: radial basis function networks, QRF: quantile regression forest, XGB: extreme gradient boosting and KG_6: Ordinary Kriging using six neighboring observations.

Methods such as GBM, RBFNs, and OK_25 had a correlation of 0.80 between nRMSE and CV, while KG, QRF, and XGB showed a correlation of 0.78. The linear mixed model showed that the interaction between interpolation methods and CV was significant, indicating different linear behaviors among methods. Using OK_25 as a reference and the average covariate value, RBFNs showed a positive nRMSE change rate with increasing CV, with nRMSE increasing by approximately 8% and 54%, respectively, for a unit increase in CV. Conversely, QRF and XGB exhibited a negative nRMSE change rate, with reductions of 8% and 5%, respectively, for the same unit increase in CV.

The computational time required for each method is presented in Figure 2. The GBM method was the fastest, with an average processing time of 1.73 seconds per map, followed by OK_6 (19.1 seconds), RBFNs (25.8 seconds), QRF (37.0 seconds), and OK_25 (38.4 seconds). XGB was the slowest method, taking an average of 382 seconds per map. Computational times were correlated with the number of observations per map. While GBM showed no significant changes in

computational time as the number of observations varied, XGB, RBFNs, and QRF exhibited a linear relationship, and OK_6 and OK_25 followed a quadratic trend (Figure 2).

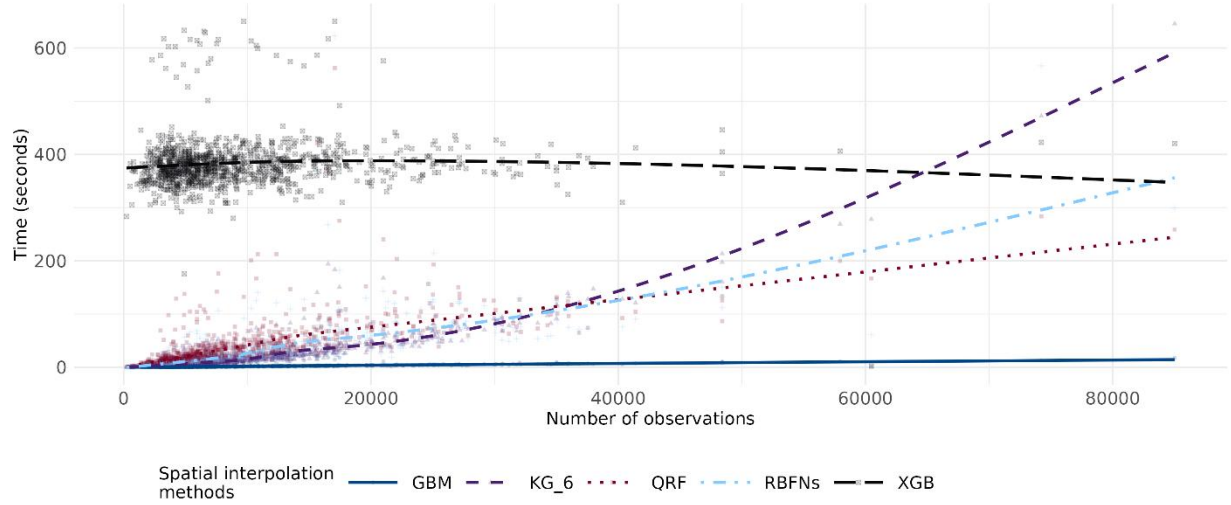

Figure 2. Relationship between the computational time of calculation and number of observations of each performance map for the evaluated methods. Generalized boosted regression model (GBM), radial basis function networks (RBFNs), quantile regression forest (QRF), extreme gradient boosting (XGB), and ordinary kriging using 6 neighboring observations (KG_6) and 25 neighboring observations (KG_25).

Across all maps, the methods showed a consistent performance pattern, with an average correlation of 0.84 between them (Figure 3). The highest average correlation was observed between KG_6 and KG_25 (0.97), while the lowest correlation was between RBFN and XGB (0.58). RBFN consistently exhibited the lowest correlations with other methods. Figure 4 displays the within-field yield prediction maps generated by the six interpolation methods. Maps created using RBFNs demonstrated greater smoothing of interpolated values, resulting in more uniform predictions compared to the other methods.

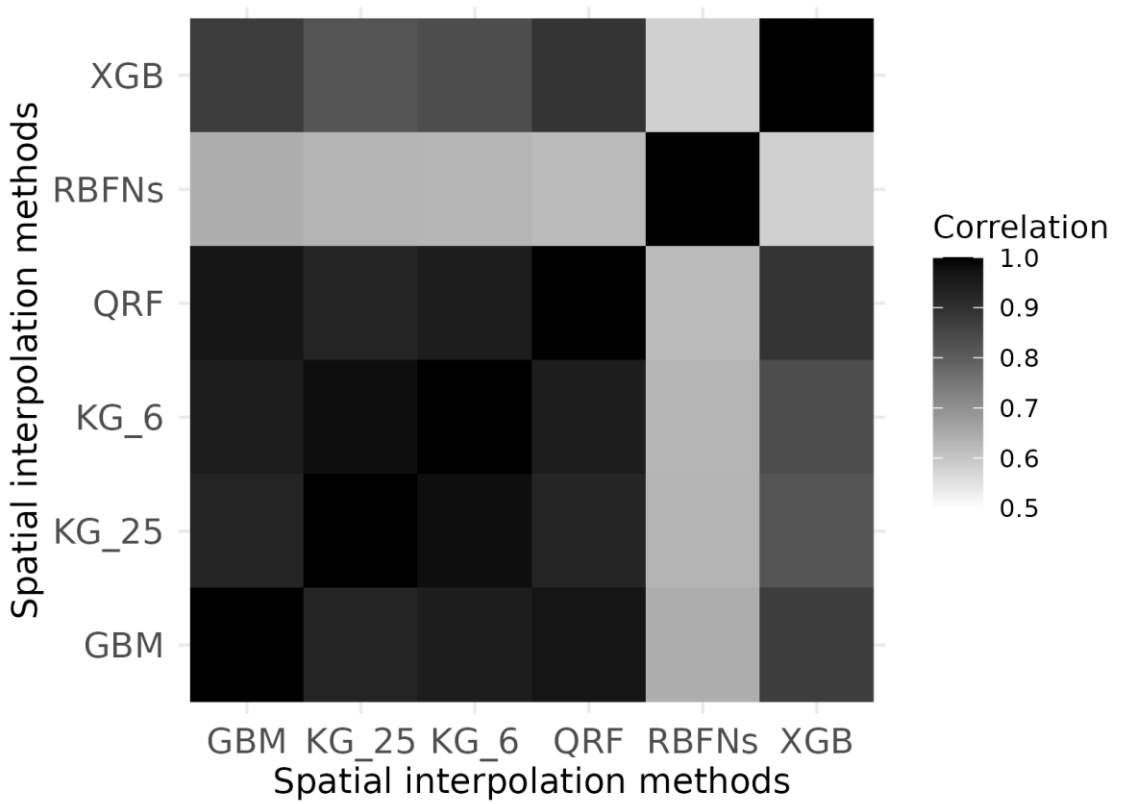

Figure 3. Correlation between the predicted values of the 1050 maps among all the methods studied. Generalized boosted regression model (GBM), radial basis function networks (RBFNs), quantile regression forest (QRF), extreme gradient boosting (XGB), and ordinary kriging using 6 neighboring observations (KG_6) and 25 neighboring observations (KG_25).

Uncertainty maps for the predictions from each interpolation method are shown in Figure 5. The two OK implementations produced similar uncertainty values, while QRF showed lower uncertainty than the other methods and followed a pattern similar to that of OK. Most methods exhibited higher uncertainty in regions with sparse or no data. RBFNs, however, showed consistently higher uncertainty values across all maps. The results of uncertainty evaluation, based on PICP plots, are illustrated in Figure 6 for three selected maps. These maps represent maize (n = 6155, CV = 20%), wheat (n = 5601, CV = 9%), and soybean (n = 3038, CV = 29%).

Overall, the methods performed better on the wheat and soybean maps, while the maize map showed greater deviation from the 1:1 line. All methods presented values above the 1:1 line, indicating over-optimistic estimates of precision or reliability. The PICP analysis revealed that for a 95% confidence level, 90% of observations fell within the expected prediction interval for all methods.

On the wheat map, the KG_6 and KG_25 methods performed similarly, with PICP values being higher than the expected confidence levels, particularly in the 95 and 99%. QRF and GBM show overestimations of the uncertainty at low confidence levels (40% and 20%), where the PIPC values were below the confidence level.

For the maize map, the QRF, RBFNs and XGB methods showed a tendency to underestimate the uncertainties, especially for confidence levels higher than 80%. The KG_6 and KG_25 methods presented PICP values that underestimated the uncertainty for lower confidence levels.

For the soybean map, the RBFN method had the highest PICP values in the widest intervals (confidence levels ≥ 95%). On the other hand, the QRF method showed a drop in confidence levels below 20%. The XGB method showed values close to the 1:1 line for the 97.5 and 99% confidence intervals.

Conversely, RBFN and XGB tend to overestimate uncertainty at lower confidence ranges, while QRF and GBM showed an intermediate performance but also overestimation tendencies at low confidence levels.

Table 3. Prediction interval coverage probability (PICP) for six interpolation methods and three crops, according to the confidence level (CI).

Crop	Method	Confidence level (CI %)									
		99	97.5	95	90	80	60	40	20	10	5

Wheat	QRF	98.0	96.2	93.8	91.2	84.7	68.1	49.8	26.6	11.8	6.3
	GBM	97.9	95.8	94.1	90.3	83.5	66.6	48.5	25.4	13.5	7.3
	KG_6	98.4	97.3	95.9	92.1	85.6	67.7	47.4	23.7	12.7	7.6
	KG_25	98.4	97.3	95.9	92.1	85.6	67.7	47.4	23.7	12.7	7.6
	RBFNs	99.5	98.6	97.1	94.6	90.5	74.1	54.3	28.9	15.5	7.0
	XGB	98.8	98.1	97.1	94.3	90.0	75.6	55.7	30.3	15.2	6.7
Maize	QRF	97.2	96.3	95.2	92.8	88.8	76.8	53.3	27.4	15.1	7.3
	GBM	97.0	95.6	94.7	93.2	89.5	79.4	59.4	32.6	16.5	8.1
	KG_6	99.0	98.5	97.6	95.8	92.4	77.9	57.4	30.7	15.2	7.5
	KG_25	99.0	98.5	97.6	95.8	92.4	77.9	57.4	30.7	15.2	7.5
	RBFNs	99.6	98.7	97.2	95.4	89.9	78.0	59.2	31.0	16.0	7.4
	XGB	99.0	98.5	97.8	96.3	93.1	80.5	60.5	33.4	17.4	8.4
Soybean	QRF	98.5	97.7	95.9	93.1	85.3	66.2	47.0	23.4	10.1	5.0
	GBM	98.4	97.0	95.1	93.4	86.5	67.2	43.9	23.9	12.5	6.8
	KG_6	98.8	97.9	96.5	92.9	86.5	68.5	45.7	22.6	11.7	5.9
	KG_25	98.8	97.9	96.5	92.9	86.5	68.5	45.7	22.6	11.7	5.9
	RBFNs	99.3	99.0	98.4	95.9	90.3	74.6	51.8	25.9	12.5	7.4
	XGB	99.3	98.5	97.9	96.2	91.1	72.6	51.3	26.6	14.7	8.1

GBM: generalized boosted regression model, RBFNs: radial basis function networks, QRF: quantile regression forest, XGB: extreme gradient boosting, KG_6: Ordinary Kriging using 6 neighboring observations and KG_25: Ordinary Kriging using 25 neighboring observations.

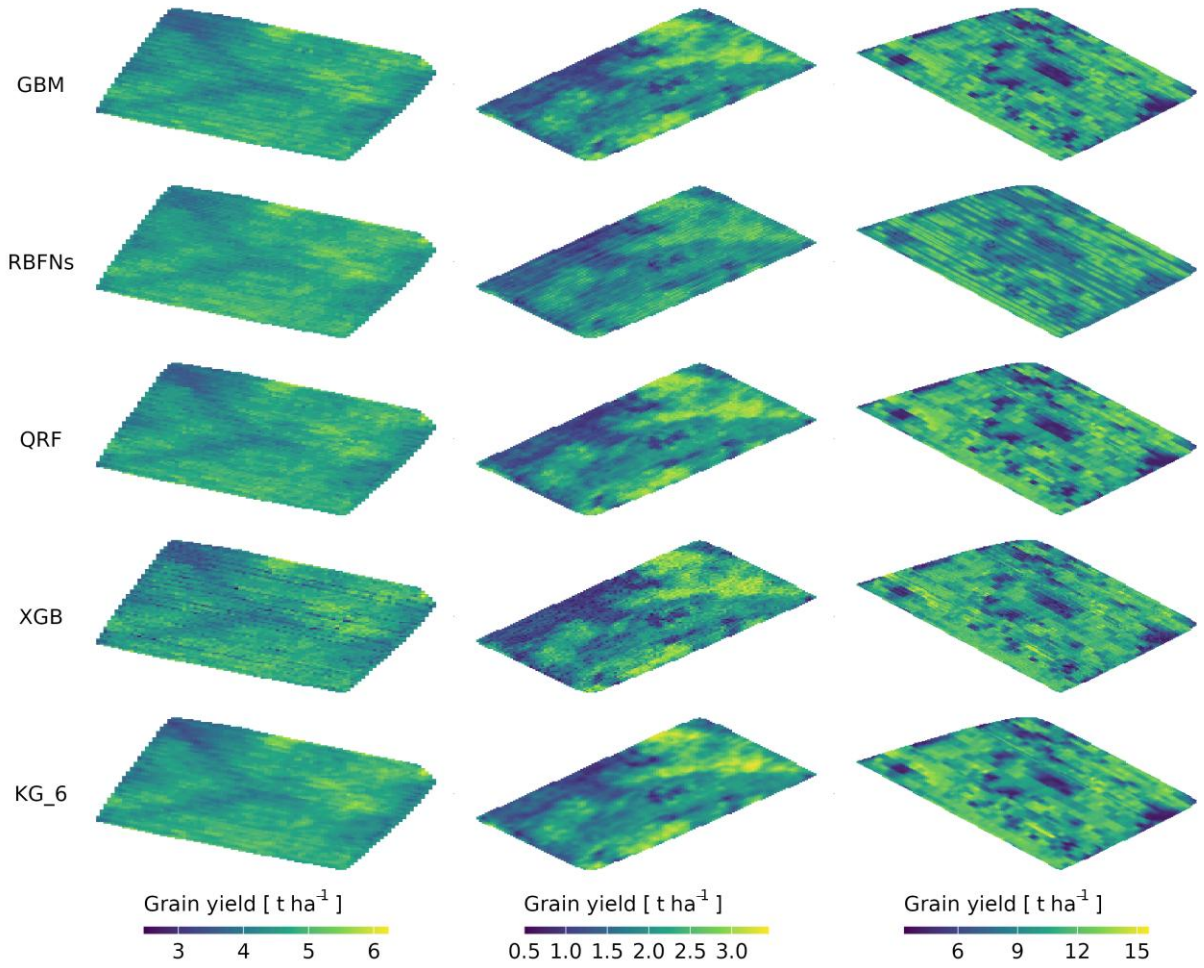

Figure 4. Within-field yield prediction maps for wheat, soybeans and corn (from left to right) with the six spatial interpolation methods evaluated: Generalized boosted regression model (GBM), radial basis function networks (RBFNs), quantile regression forest (QRF), extreme gradient boosting (XGB) and ordinary kriging using 6 neighboring observations (KG_6).

354

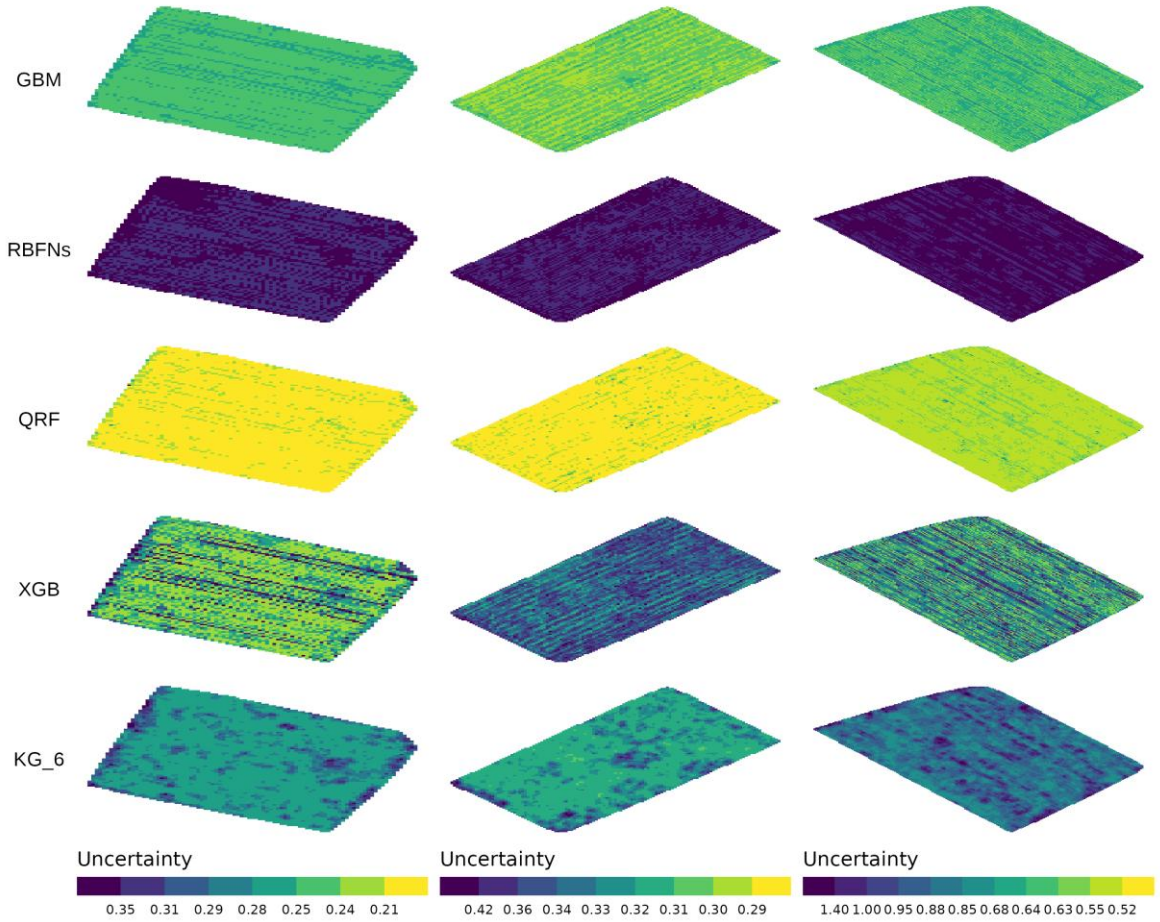

Figure 5. Uncertainty maps of within-field yield prediction for wheat, soybeans and corn (from left to right) with the six spatial interpolation methods evaluated. Generalized boosted regression model (GBM), radial basis function networks (RBFNs), quantile regression forest (QRF), extreme gradient boosting (XGB) and ordinary kriging using 6 neighboring observations (KG_6).

355

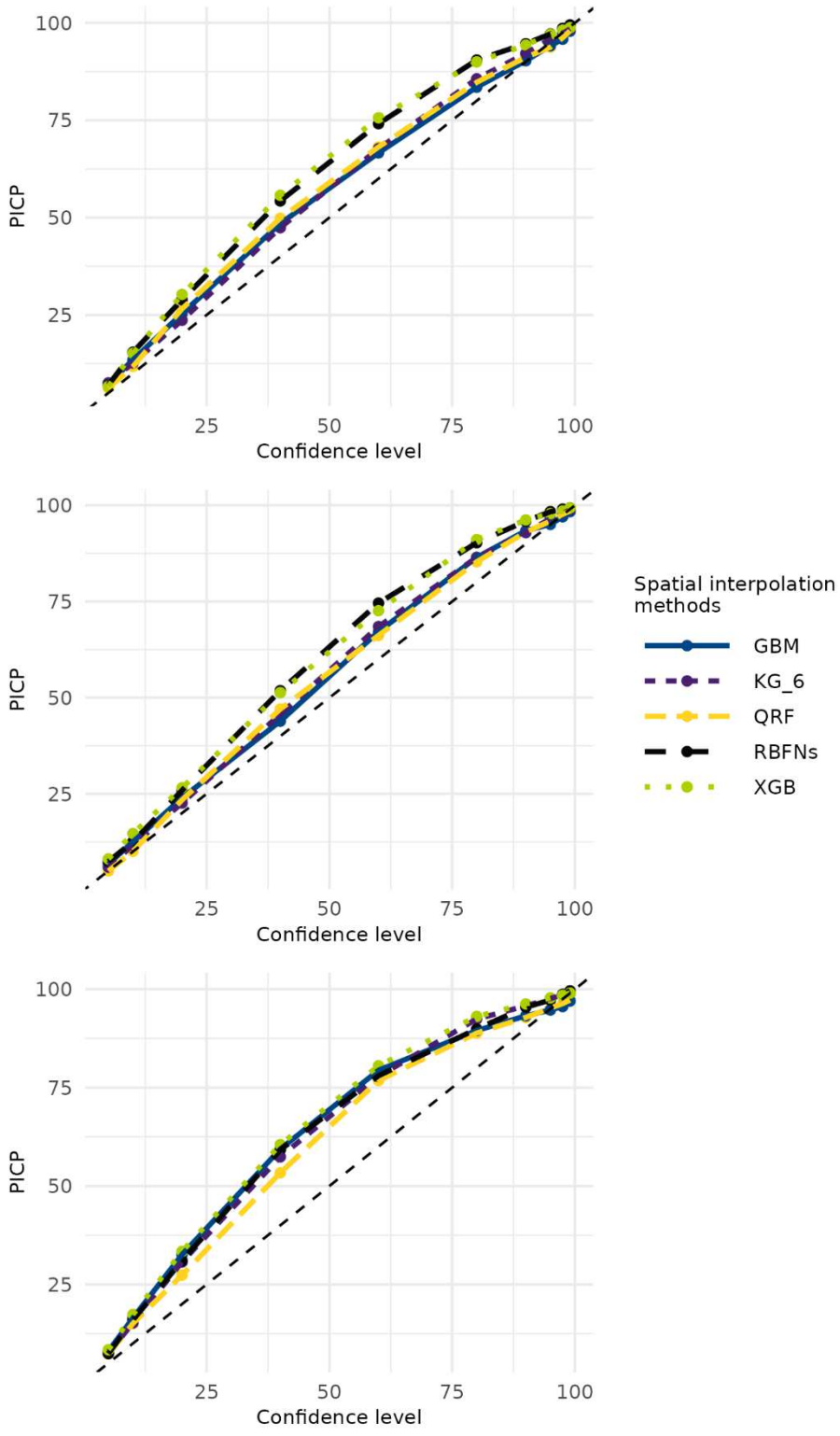

Figure 6. PICP confidence Plot for validation-based performance by bootstrapping approach for each method in wheat, soybean, and corn (top to bottom). The 1:1 black dotted line represents the desired outcome.

4 DISCUSSION

Mapping the spatial variability of crop yields is a critical step for implementing site-specific management practices in PA. Various techniques have been proposed to explore spatial distribution patterns and interpolation methods for within-field yield mapping, including classical geostatistical methods and, more recently, ML algorithms. This study compared different methods for mapping the spatial variability of crop yield. Predictive capability was evaluated using cross-validation processes and nRMSE as an error indicator, as well as the estimation of the predictions' standard error and PICP for assessing predictive uncertainty. Traditional methods like kriging have been widely used for mapping spatial variability and addressing data discontinuity (Kim et al., 2022), especially in datasets of smaller dimensions. However, geostatistical methods had a significant increase in computational time with larger datasets. In this study, kriging computational time increased exponentially with the number of observations. ML methods, except for XGB, exhibited lower computational times than OK, particularly for datasets exceeding 20,000 observations. This highlights the capacity of ML algorithms to handle large datasets effectively (Ahmad et al., 2022; Hristopulos et al., 2020). Among ML algorithms, GBM demonstrated the shortest computational time, followed by QRF (Figure 2).

Regarding predictive accuracy, QRF achieved the lowest nRMSE among ML methods, demonstrating superior interpolation precision (Table 1). In contrast, RBFN exhibited the highest prediction error. One possible explanation for RBFN's lower performance is its susceptibility to overfitting when working with tabular data (i.e., datasets organized in rows and columns of variables), where high dimensionality or over-parameterization can hinder model accuracy (Kim et al., 2022). Conversely, tree-based models such as GBM and XGB are better suited for high-dimensional tabular data, balancing bias and variance to avoid overfitting (Breiman, 2001;

Lundberg et al., 2018). In this study, both GBM and XGB performed well but were outperformed by QRF (Table 2). ML methods generally require hyperparameter tuning to achieve optimal performance (Probst et al., 2019). As noted by Kim et al., (2022), achieving effective neural network performance involves iteratively refining parameters such as the number of layers, hidden nodes, and other hyperparameters. In this study, default hyperparameter settings provided by software implementations were used, which may have influenced the results. While geostatistical methods require complex statistical assumptions, such as data stationarity, normal distribution, and linear relationships between variables, these assumptions can complicate the identification of a robust model (Córdoba and Balzarini, 2021). On the other hand, ML algorithms offer flexibility in dealing with non-linear and non-stationary data. The similarities observed between XGB and GBM may be attributed to their analogous framework, though XGB benefits from regularized boosting and parallel processing, making it more efficient and less prone to overfitting than GBM (Bentéjac et al., 2021; Chen & Guestrin, 2016). Previous studies employing hyperparameter optimization via grid search with stratified cross-validation (GridSearchCV) suggested that XGB's predictive efficiency often exceeds that of other ML models and OK (Ibrahim et al., 2022). Here, GridSearchCV systematically evaluated combinations of model hyperparameters using cross-validation to identify the optimal settings.

Bootstrap methods are used to estimate uncertainty in ML-based interpolations, but they are computationally demanding (Kasraei et al., 2021). OK provided an advantage in this regard, since it calculates prediction variances inherently. Among the ML methods, QRF is notable for its built-in uncertainty quantification mechanism (Kasraei et al., 2021). In this study, uncertainty estimates were derived through resampling methods. QRF displayed lower uncertainty values, whereas GBM and XGB presented similar uncertainty levels to those of OK using this methodology.

Notably, QRF offers the ability to quantify uncertainty directly during the application of the method, avoiding the use of computationally intensive approaches like bootstrapping. For large datasets, resampling increases computational demand significantly. QRF's inherent uncertainty quantification, based on the same tree forest used for property estimation, aligns with kriging's approach as an "elegant way of deriving uncertainty models for spatially distributed properties" (Heuvelink, 2013). The findings support the superiority of QRF in terms of both predictive accuracy and uncertainty quantification, demonstrating its potential as a robust alternative to traditional geostatistical methods and other ML algorithms. Future research could focus on optimizing hyperparameters for all methods to maximize their predictive capabilities as well as on exploring hybrid approaches to further enhance accuracy and computational efficiency.

5 CONCLUSIONS

Machine learning methods, particularly QRF, GBM, and XGB, consistent statistical performance, successfully handling large datasets and enhancing the precision of spatial interpolation, with QRF achieving the lowest prediction error and enhancing spatial interpolation precision by reducing error up to 13% compared to OK. Although QRF requires careful hyperparameter tuning, its predictive efficiency highlights its potential as a powerful tool for mapping within-field yield variability. Furthermore, QRF offers the advantage of generating high-quality uncertainty maps, enabling a robust assessment of prediction reliability across yield maps. These findings underscore the promise of ML methods, especially QRF, in advancing PA practices and their adoption.

ACKNOWLEDGMENTS

We thank the National University of Córdoba (UNC), the National Scientific, Argentine National Scientific and Technological Promotion Agency (ANPCyT-PICT 2018 - 3321), Ministry of Science and Technology of Cordoba province, Secretary for Science and Technology of National University of Córdoba (UNC), and the National Scientific and Technical Research Council (CONICET), for supporting this research.

REFERENCES

Ahmad, W., Muhammad, K., Glass, H. J., Chatterjee, S., Khan, A., & Hussain, A. (2022). Novel MLR-RF-Based Geospatial Techniques: A Comparison with OK. *ISPRS International Journal of Geo-Information*, *11*(7). https://doi.org/10.3390/ijgi11070371

Al-Mudhafar, W. J., Abbas, M. A., & Wood, D. A. (2022). Performance evaluation of boosting machine learning algorithms for lithofacies classification in heterogeneous carbonate reservoirs. *Marine and Petroleum Geology*, *145*. https://doi.org/10.1016/J.MARPETGEO.2022.105886

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*(3), 1937–1967. https://doi.org/10.1007/S10462-020-09896-5/TABLES/12

Bergmeir, C. N., & Benítez Sánchez, J. M. (2012). *Neural networks in R using the Stuttgart neural network simulator: RSNNS*.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bronowicka-Mielniczuk, U., & Mielniczuk, J. (2023). New indices to quantify patterns of relative errors produced by spatial interpolation models – A comparative study by modelling soil properties. *Ecological Indicators*, *154*, 110551. https://doi.org/https://doi.org/10.1016/j.ecolind.2023.110551

Burdett, H., & Wellen, C. (2022). Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precision Agriculture*, *23*(5), 1553–1574. https://doi.org/10.1007/s11119-022-09897-0

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., & Khotilovich, V. (2019). Package 'xgboost.' *R Version*, *90*(1–66), 40.

Cordoba, M., & Balzarini, M. (2021). A random forest-based algorithm for data-intensive spatial

interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, *184*, 106094.

Córdoba, M., & Balzarini, M. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, *184*. https://doi.org/10.1016/j.compag.2021.106094

Dang, S., Peng, L., Zhao, J., Li, J., & Kong, Z. (2022). A Quantile Regression Random Forest-Based Short-Term Load Probabilistic Forecasting Method. In *Energies* (Vol. 15, Issue 2). https://doi.org/10.3390/en15020663

Giraldo, R., Leiva, V., & Castro, C. (2023). An overview of kriging and cokriging predictors for functional random fields. *Mathematics*, *11*(15), 3425.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, *340*(November 2018), 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Greenwell, B., Boehmke, B., Cunningham, J., Developers, G. B. M., & Greenwell, M. B. (2019). Package 'gbm.' *R Package Version*, *2*(5).

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learnin. *Springer Series in Statistics*, 33.

Hiemstra, P., & Hiemstra, M. P. (2013). Package 'automap.' *Compare*, *105*, 10.

Hristopulos, D. T., Pavlides, A., Agou, V. D., & Gkafa, P. (2020). Stochastic Local Interaction Model: Geostatistics without Kriging. *Mathematical Geosciences*, *53*(8), 1907–1949. https://doi.org/10.1007/s11004-021-09957-7

Ibrahim, B., Majeed, F., Ewusi, A., & Ahenkorah, I. (2022). Residual geochemical gold grade prediction using extreme gradient boosting. *Environmental Challenges*, *6*, 100421. https://doi.org/https://doi.org/10.1016/j.envc.2021.100421

Imanian, H., Shirkhani, H., Mohammadian, A., Hiedra Cobo, J., & Payeur, P. (2023). Spatial Interpolation of Soil Temperature and Water Content in the Land-Water Interface Using Artificial Intelligence. *Water (Switzerland)*, *15*(3). https://doi.org/10.3390/w15030473

Karapetsas, N., Alexandridis, T. K., Bilas, G., Munnaf, M. A., Guerrero, A. P., Calera, M., Osann, A., Gobin, A., Rezník, T., Moshou, D., & Mouazen, A. M. (2022). Mapping Soil Properties with Fixed Rank Kriging of Proximally Sensed Soil Data Fused with Sentinel-2 Biophysical Parameter. In *Remote Sensing* (Vol. 14, Issue 7). https://doi.org/10.3390/rs14071639

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609–629. https://doi.org/https://doi.org/10.1016/j.ins.2021.11.036

Kasraei, B., Heung, B., Saurette, D. D., Schmidt, M. G., Bulmer, C. E., & Bethel, W. (2021). Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. *Environmental Modelling & Software*, *144*, 105139. https://doi.org/https://doi.org/10.1016/j.envsoft.2021.105139

Kerry, R., Oliver, M. a, & Frogbrook, Z. L. (2010). Geostatistical Applications for Precision

Agriculture. *Precision Agriculture*, 305–312. https://doi.org/10.1007/978-90-481-9133-8

Kim, J., Lee, Y., Lee, M.-H., & Hong, S.-Y. (2022). A Comparative Study of Machine Learning and Spatial Interpolation Methods for Predicting House Prices. In *Sustainability* (Vol. 14, Issue 15). https://doi.org/10.3390/su14159056

Kirkwood, C., Cave, M., Beamish, D., Grebby, S., & Ferreira, A. (2016). A machine learning approach to geochemical mapping. *Journal of Geochemical Exploration*, *167*, 49–61. https://doi.org/10.1016/J.GEXPLO.2016.05.003

Koutsos, T. M., Menexes, G. C., Eleftherohorinos, I. G., & Alexandridis, T. K. (2023). Using Block Kriging as a Spatial Smooth Interpolator to Address Missing Values and Reduce Variability in Maize Field Yield Data. In *Agronomy* (Vol. 13, Issue 7). https://doi.org/10.3390/agronomy13071685

Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, *38*(4), 963. https://doi.org/10.2307/2529876

Lin, L. I.-K. (1992). Assay Validation Using the Concordance Correlation Coefficient. *Biometrics*, *48*(2), 599–604. https://doi.org/10.2307/2532314

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles*. *2*.

Max Kuhn. (2021). *Package "caret" Title Classification and Regression Training*.

Minasny, B., & McBratney, A. B. (2005). The Matérn function as a general model for soil variograms. *Geoderma*, *128*(3), 192–207. https://doi.org/https://doi.org/10.1016/j.geoderma.2005.04.003

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. In *Catena* (Vol. 113, pp. 56–69). Elsevier. https://doi.org/10.1016/j.catena.2013.09.006

Oliver, M., Catena, R. W.-, & 2014, undefined. (n.d.). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Elsevier*.

Pebesma, E., & Graeler, B. (2012). Spatial and spatio-temporal geostatistical modelling, prediction and simulation. *The Comprehensive R Archive Network (CRAN)*.

Piepho, H. (2019). A coefficient of determination (R2) for generalized linear mixed models. *Biometrical Journal*, *61*(4), 860–872.

Probst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, *20*, 1–32.

R Core Team. (2024). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*.

Ridgeway, G. (1999). The state of boosting. *Computing Science and Statistics*, *31*, 172–181.

Salem, H. M., Schott, L. R., Piaskowski, J., Chapagain, A., Yost, J. L., Brooks, E., Kahl, K., & Johnson-Maynard, J. (2024). Evaluating Intra-Field Spatial Variability for Nutrient Management Zone Delineation through Geospatial Techniques and Multivariate Analysis. *Sustainability (Switzerland)*, *16*(2), 645. https://doi.org/10.3390/SU16020645/S1

Schabenberger, O., & Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. Chapman and Hall/CRC.

Schmidinger, J., & Heuvelink, G. B. M. (2023). Validation of uncertainty predictions in digital soil mapping. *Geoderma*, *437*, 116585. https://doi.org/https://doi.org/10.1016/j.geoderma.2023.116585

Sekulić, A., Kilibarda, M., Heuvelink, G. B. M., Nikolić, M., & Bajat, B. (2020). Random Forest Spatial Interpolation. *Remote Sensing 2020, Vol. 12, Page 1687*, *12*(10), 1687. https://doi.org/10.3390/RS12101687

Sergeev, A. P., Buevich, A. G., Baglaeva, E. M., & Shichkin, A. V. (2019). Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. *CATENA*, *174*, 425–435. https://doi.org/https://doi.org/10.1016/j.catena.2018.11.037

Siabato, W., Guzmán-Manrique, J., Siabato, W., & Guzmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos de Geografía: Revista Colombiana de Geografía*, *28*(1), 1–22. https://doi.org/10.15446/RCDG.V28N1.76919

Silva, C. de O. F., Manzione, R. L., & Oliveira, S. R. de M. (2023). Exploring 20-year applications of geostatistics in precision agriculture in Brazil: what's next? *Precision Agriculture*, *24*(6), 2293–2326. https://doi.org/10.1007/s11119-023-10041-9

Taylor, J. A., Acevedo-Opazo, C., Ojeda, H., & Tisseyre, B. (2010). Identification and significance of sources of spatial variation in grapevine water status. *Australian Journal of Grape and Wine Research*, *16*(1), 218–226. https://doi.org/10.1111/j.1755-0238.2009.00066.x

Wadoux, A. M. J. C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. In *Earth-Science Reviews* (Vol. 210, p. 103359). Elsevier. https://doi.org/10.1016/j.earscirev.2020.103359

Wright, M., arXiv:1508.04409, A. Z. preprint, & 2015, undefined. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Arxiv.Org*, *77*(1). https://doi.org/10.18637/jss.v077.i01