Journal of Environmental Management 289 (2021) 112509

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: http://www.elsevier.com/locate/jenvman

Research article

A spatially based quantile regression forest model for mapping rural land values

Mariano Córdoba [a,b,*], Juan Pablo Carranza [c], Mario Piumetto [d,e], Federico Monzani [d], Mónica Balzarini [a,b]

[a] Universidad Nacional de Córdoba, Facultad de Ciencias Agropecuarias, Cátedra de Estadística y Biometría, Córdoba, Argentina
[b] Unidad de Fitopatología y Modelización Agrícola (UFyMA), INTA – CONICET, Córdoba, Argentina
[c] Universidad Nacional de Córdoba, Instituto de Investigación y Formación en Administración Pública (IIFAP), Córdoba, Argentina
[d] Infraestructura de Datos Espaciales de La Provincia de Córdoba (IDECOR), Argentina
[e] Universidad Nacional de Córdoba, Facultad de Ciencias Exactas, Físicas y Naturales, Centro de Estudios Territoriales (CET), Córdoba, Argentina

ARTICLE INFO

Keywords:
Mass appraisal
Machine learning
Spatial autocorrelation
Prediction uncertainty

ABSTRACT

Rural land valuation plays an important role in the development of land use policies for agricultural purposes. The advance of computational software and machine learning methods has enhanced mass appraisal methodologies for modeling and predicting economic values. New machine learning methods, like tree-based regression models, have been proposed as an alternative to linear regression to predict economic values from ancillary variables, since these algorithms are able to handle non-normality and non-linearity in the data. However, regression trees are commonly estimated assuming independent rather than spatially correlated data. This study aims to build a tree-based regression model that will help to tackle methodological problems related to the determination of prices of rural lands. The Quantile Regression Forest (QRF) algorithm was used to provide a regression model to predict and assess the uncertainty associated with model-derived predictions. However, the classical QRF ignores the autocorrelation underlying spatialized land values. The objective of this work was to develop, implement, and evaluate a spatial version of QRF, named sQRF, for computer-assisted mass appraisal of rural land values accounting for information from neighboring sites. We compared predictions of land values from sQRF with those obtained from spatial random forest, kriging regression, and linear regression models. sQRF performed well in predicting rural land values; indeed, it performed better than multiple linear regression. An important feature of sQRF is its ability to produce a direct uncertainty measure to assess the goodness of the predictions. Land values reflect a complex mix of agricultural returns, localization, and access to markets, which can be predicted from ancillary environmental variables. Good predictive models are essential to determine land values for multiple purposes including territorial taxation.

1. Introduction

Land valuation plays an important role in the development of land use policies for many purposes. Knowledge and permanent monitoring of land value in the market are key elements in the design of soil management policies and overall land planning and contribute to the following aspects, among others: i) improved recouping of public investments by the State; ii) increased control of speculative practices of land used, which increase of land value; iii) identification of zones with higher or lower value and the variables involved in such valuation, with the aim of defining government interventions that promote more equitable land development and more efficient funding. Land values reflect a complex mix of economical returns, localization, access to markets, and other site-specific features (Choumert et al., 2014). Therefore, land unit values (LUVs) can be predicted from an ancillary set of environmental or explanatory variables. The 2030 agenda for sustainable development, known as Global Goals, linked the importance of improved land use planning, administration, and management (Bencure et al., 2019); however, defining a precise and perfect valuation model remains a difficult task due to variations in the related factors (Sesli, 2015). Particularly, there is a great interest in assigning values to rural land because rural development is based on the strengthening of land

* Corresponding author.
E-mail address: marianoacba@agro.unc.edu.ar (M. Córdoba).

https://doi.org/10.1016/j.jenvman.2021.112509
Received 2 November 2020; Received in revised form 16 February 2021; Accepted 26 March 2021
Available online 6 April 2021
0301-4797/© 2021 Elsevier Ltd. All rights reserved.

markets. However, outdated cadastral data, financial speculation and inflationary processes can be major obstacles for developing precise land valuations, hindering the use of price formation mechanisms from the traditional market (Caballer, 2008). New predictive models (Yacim and Boshoff, 2018) appear as suitable methods for mass appraisal, even in cases in which the implementation of conventional valuation methods is cumbersome.

The statistical approach to mass appraisal allows price assessment using sampling data and standard methods to predict prices of unsampled properties from ancillary variables. Computer-assisted mass appraisal has become popular as a partially automated valuation model (McCluskey et al., 2013; Piumetto et al., 2019; Zhang et al., 2015). This type of models can estimate land value based on analysis of location, market conditions, and real estate characteristics used as predictors (IAAO, 2017). The linear regression model (LR) is the most popular statistical model used to develop a predictive equation for use in computer-assisted mass appraisal valuations (Demetriou, 2017). The classical LR model relating property values to several site-specific attributes is simple and interpretation is straightforward; however, its high computational efficiency decreases in large samples of spatialized data. While the LR model assumes that relationships between the value of properties and explanatory variables are linear, relationships are usually complex and non-linear. In the last years, other predictive models built from machine learning methods have arisen as an alternative approach to computer-assisted mass appraisal, since they can successfully handle non-normality and non-linearity in the sampled data. One of the most widely used machine learning methods in land valuation is the artificial neural network (Antipov and Pokryshevskaya, 2012). The accuracy of a neural network is comparable and sometimes better than the LR model (Demetriou, 2017). Kontrimas and Verikas (2011) showed that several machine learning methods performed better in the assignment of values to land than the official real estate methods. Using machine learning algorithms, Antipov and Pokryshevskaya (2012) and Čeh et al. (2018) achieved a significant improvement in the accuracy and fairness of the land valuation results. Within this framework, regression models can be adjusted from a tree-based regression algorithm (Breiman et al., 1984) or its extension to enhance prediction from resampling, such as the Random Forest (RF) (Breiman, 2001) as well as Quantile Random Forest (QRF) (Meinshausen, 2006). The machine learning methods powered by resampling, such as RF or QRF, are considered more resistant to overfitting and more robust to noise in the data than regular regression tree models. Some of this type of machine learning models have been successfully used for mass appraisal in residential real estate (Wang and Li, 2019). However, RF and QRF do not account for spatiality in the data. Rural property values are usually spatialized and exhibit a stronger positive autocorrelation (Uberti et al., 2018). The presence of positive autocorrelation in rural land values leads that nearby locations have more similar land prices than distant locations. Therefore, spatial models, i.e. models that account for the spatial correlation, are required since they will use additional information on neighborhood/environmental variables that would otherwise not have been included in an analysis using models that assume independent data.

Different approaches are used in the economic literature to address the issue of spatial correlations. Spatial econometrics models (e.g. spatial lag model, spatial error model), and geographically weighted regressions (Bidanset and Lombard, 2014; Georgiadis, 2018; Jahanshiri et al., 2011; Uberti et al., 2018; Zhang et al., 2015) are common examples of spatially restricted predictive models. However, there is a gap in the literature about spatial data modeling with machine learning methods. When predicting the value of interest at a given site, commonly used machine learning methods ignore influences (spatial autocorrelation) of observed neighboring data. Recently, a spatially tree-based model has been proposed by adding kriged residuals to the model prediction located in a spatial domain (Guo et al., 2015; Hengl et al., 2015). The spatially restricted tree-based regression models are quite similar to the popular regression kriging method (RK) used in geostatistics in the manner of handling spatial correlations (Hengl et al., 2007). Once the predictive model has been developed from a sample selected for model calibration, site-specific predictions can be obtained for lands not included in the sample used to fit the model. Therefore, the fitted model is then used to produce continuous land values useful for digital economy (Simões et al., 2020). It is clear that each predictive modeling technique has specific advantages and disadvantages, with predictive accuracy being considered the fundamental component; therefore, each technique should be measured using standard predictive performance measures and benchmarked relative to overall predictive accuracy (McCluskey et al., 2013). The objectives of this paper were to introduce and comparatively assess a new spatially restricted machine learning approach for predicting rural LUVs. Its predictive accuracy was measured against recognized statistical tools to fit predictive models.

The fitted model will be used to produce a digital map of LUVs. The variability map is a simplified representation of a complex and partially unknown spatial process defining the land market values. Therefore, a digital map generated by model-based predictions of land values has an irreducible uncertainty that must be considered during its interpretation. Geostatistical models such as RK are, in essence, adequate for providing uncertainty estimates from theory (Goovaerts, 2001). The most common error measurement in RK is the kriging variance (Oliver and Webster, 2014). For the machine learning methods used to build predictive models, formal quantification of uncertainty is quite novel (Szatmári and Pásztor, 2019), with QRF being the recommended method to empirically derive prediction error measurements in scenarios of independent data (Hengl et al., 2018). This paper presents the development, implementation, and evaluation of a spatial version of QRF applied to visualize the spatial variability of rural land values and its uncertainties. The spatial tree-based model allows a further understanding of the relationship between site-specific variables and land prices. The modeling results may allow us to redesign mechanisms for taxation of the rural lands. To the best of our knowledge, the QRF model has still not been applied as predictor of LUVs from regionalized data. As a result, this study is expected to add to the extant literature on methodologies for digital mapping of rural land values.

2. Materials and methods

2.1. Case study

Data from a rural land re-valuation conducted in Córdoba, Argentina, was used as a case study. The territory of Córdoba province covers 165,321 km^2 and is divided in approximately 200,000 rural plots. The relief of the province is characterized by two main mountain ranges to the west (sierras) and valleys or plains of fertile soil to the east and south. The high topographic, climatic and edaphological variability (Giannini Kurina et al., 2018) generates different soil aptitudes and land use capabilities (33% of lands suitable for agriculture and 17% of lands regarded as marginal for agriculture). For mass appraisal, data were generated within the framework of the "real estate territorial study" project conducted by the Government of the Province of Córdoba, Argentina. A Real Estate Market Observatory was implemented (OMI - http://omi.mapascordoba.gob.ar/); a sample of 3480 rural land market data encompassing the 2018–2019 period was systematized.

2.2. Data preprocessing and explanatory variables

Data were cleaned using global and spatial outlier analyses. For the latter analysis, each observation was compared with the market values surveyed within a proximity radius of 5000 m and those observations showing marked disparity from their neighbors (spatial outliers) were identified using the local Moran index. The explanatory variables used to perform the regression models, most of them continuous variables, are included in Table 1. Native forest areas and flood-prone areas, which

Table 1
Explanatory variables used to build a mass appraisal model in Córdoba, Argentina.

Type	Name	Source
Cadastral	Number of plots in the surroundings (10 km)	Cadastral Agency of Córdoba province
	Minimum plot area in the surroundings (10 km)	
	Maximum plot area in the surrounding (10 km)	
	Mean plot area in the surroundings (10 km)	
	Previous valuation per zone (1994)	
	% Constructed area	
Legislation	Edapho-climatic zoning	IDECOR
	Environmentally protected Area (Act Nº 9164)	Association of Agronomic Engineers of Córdoba
	Zones of appraisal value	IDECOR
	% Area according to categories of Land planning of Native Forests (OTBN Act 9814)	Act of Land Planning of Native Forest and regulation of exotic forests of the Province of Córdoba
	% Area according to land fragmentation	IDECOR
Soil and Vegetation	Soil productivity index	National Institute of Agricultural Technology/ Government of the province of Córdoba
	% Area according to land use capacity	
	% Area according to the land cover	IDECOR
	Soil pH	
	Nitrogen content	
	Phosphorus content	
	Potassium content	
	Cation exchange capacity	
	Organic matter content	
	% soil clay	
	Mean NDVI (historical mean 2008–2017)	Product MCD3A4 of MODIS sensor
Topography	Mean slope (%)	SRTM 90 m
	Mean elevation (m asl)	
Hydrology	Distance to the main rivers	Government of the province of Córdoba
	Depth to groundwater	
	Surface water occurrence	Global Surface Water - 1984–2018
Climate	Mean historical Palmer drought severity Index	TerraClimate-1960-2018
	Mean historical hydric deficit	TerraClimate-1960-2019
	Mean annual cumulative precipitation	WorldClim Version 2 - 1970–2000
	Mean annual maximum temperature	
	Mean annual minimum temperature	
	Mean annual temperature	
	Mean cumulative solar radiation	
	Mean monthly cumulative evapotranspiration of the series	Product MOD16A2.006: Terra Net Evapotranspiration - 2001–2018
Location	Distance to urban centers with more than 2000 inhabitants	IDECOR
	Distance to paved road network	
	Distance to regionally important locality	
	Distance to power grid	
	Distance to locality with grain storage facilities	
	Distance to a port	
Economic	Agricultural lease in the region	Economic reports -Bolsa de Cereales de Córdoba
	Regional soybean yield	
	Regional corn yield	

were detected from historical precipitation and flood series (Pekel et al., 2016), were analyzed taking special considerations due to the impact of both conditions on tax policy and the associated fiscal valuation.

The spatial pattern of each variable was modeled to obtain a fine grid (500 × 500 m) and adapt information layers from different sources and formats. Each soil variable was processed using a spatial structure analysis by universal kriging. Experimental semivariograms were calculated and suitable semivariogram models were fitted using WLS (Weighted Least Squares) (Oliver and Webster, 2014). The climate covariates were brought to the same spatial resolution by resampling using the cubic convolution method (Keys, 1981). The multiple spatial data sources used were integrated through the spatial data infrastructure of Córdoba province (IDECOR), which facilitates the acquisition, processing, storage, upgrading, and publication of spatial data, contributing to the management of public policies related to the provincial territory (Piumetto, 2020). The results of this study were applied to the local tax policy and are freely available for other uses (www.mapascordoba.gob.ar).

Variable selection was performed using the Boruta algorithm (Kursa and Rudnicki, 2010), which implements a wrapper approach based on the methodology implemented here with the land value as dependent variable. The main rationale of this approach is to compare the importance of the real predictor variables with those of random, so-called shadow, variables using several runs of modeling and statistical testing. In each run, the set of predictor variables is doubled by adding a copy of each variable. The values of shadow variables are generated by permuting the original values across observations. An RF was trained on the extended data set and the variable importance values were collected. A statistical test was performed to compare the importance of each real variable with the maximum value of all the shadow variables. Variables with significantly greater or smaller importance values were declared important or unimportant, respectively. All unimportant variables and shadow variables were removed, and the previous steps were repeated until all variables were classified or a pre-specified number of runs has been performed. This process was applied using a p-value of 0.01 with multiplicity adjustment following Bonferroni criterion. The default value of 300 was used for the maximal number of importance source runs. The Boruta algorithm was run using "Boruta" package of R.4.0 (R Core Team, 2020) software.

2.3. Developing a spatial version of the QRF model

The QRF algorithm is an ensemble of several regression trees (Breiman, 2001). The goal of a regression tree is to understand the relationship between a dependent variable (Y) and a set $p-$ dimensional of explanatory variables (X). The regression tree algorithm employs recursive data partitions according to selected X variables to create groups that are homogeneous in Y. The binary partition continues until further splitting adds no further improvement of Y heterogeneity in each formed group of observations. The findings are graphically visualized as a binary tree (Fig. 1).

The prediction in a tree-based model, given covariate $X = x$ at the new site s_0, is a weighted average of the original observations $Y_i, i = 1,\ldots, n$,

$$\widehat{Y}(s_0) = \sum_{i=1}^{n} w_i(x,\theta) Y_i$$

where $w_i(x,\theta)$ is the weight given by a positive constant if observation x_i is part of the subset, and θ indicates the considered covariate variable at each node. In our work, the group mean was regarded as the predicted value at each site.

The RF algorithm (Breiman, 2001) ensembles the regression tree models formed from several bootstrap samples. A randomly selected subset of covariates is chosen for each random sample (the mtry parameter is the number of covariates used to build each of the random

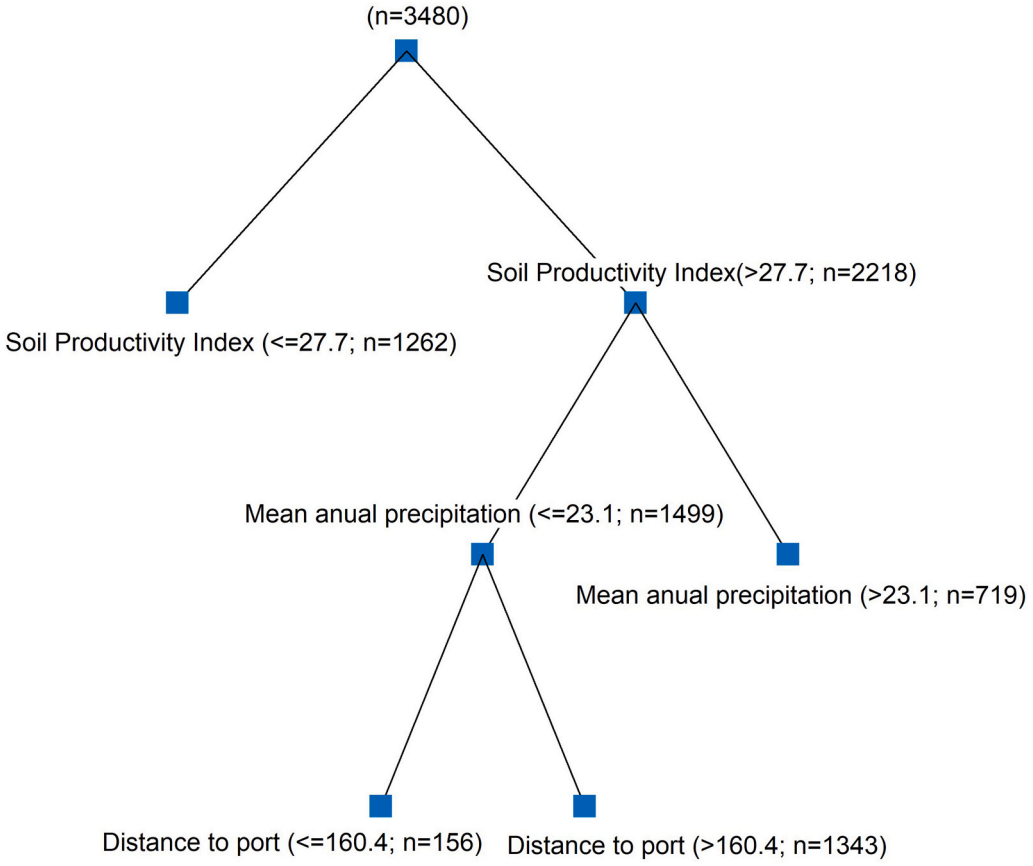

Fig. 1. Schematic visualization of regression tree results. LUV (Land unit value) is the dependent variable and soil productivity index, mean annual precipitation and distance to port are explanatory variables. The variable soil productivity index performed the first split of data, forming groups of sites with similar LUVs within groups. Therefore, it was the variable with highest predictive ability. Three nodes were formed to generate six groups of observations with relatively homogeneous LUVs. These three variables will be used to predict LUVs at a new site.

regression trees). In RF, averaging over all tree-based models provides better predictions than using a single regression tree. The RF algorithm gives an accurate approximation of the conditional mean $E(Y|X = x)$, but also provides information of the full conditional distribution of the dependent variable. Therefore, conditional quantiles can be inferred with QRF (Meinshausen, 2006), a generalization of RF. The conditional distribution of Y for $X = x$ is defined as $F(y|X = x) = P(Y \leq y|X = x)$. For estimation of $F(y|X = x)$, a weighted empirical cumulative distribution function is considered:

$$\widehat{F}(y|X=x) = \sum_{i=1}^{n} w_i(x,\theta) Y_{\{Y \leq y\}}$$

The tree-based model generated here using QRF was developed as in the standard RF. While in RF for each node in each tree, only the mean of the observations that fall into this node is kept, in the QRF method, the values of all observations in each node are kept. The full set of observations in the node was used to obtain the quantiles that were then used to build a prediction interval as a measurement of the prediction uncertainty.

The QRF model was trained, tuned, and cross-validated using the "caret" and "gstat" packages of R software (R Core Team, 2020). The "train" function within the "caret" package was used to tune candidate values. The "mtry" parameter was optimized in the interval $[2;p]$. The minimum number of terminal nodes and the number of trees was held constant in 5 and 1000, respectively. The performance of the QRF algorithm under a specific tune parameter was evaluated using a k-fold cross-validation algorithm (k = 10). In this procedure, the original sample was randomly partitioned into k equal-sized subsamples. The prediction model was trained on *k-1* subsamples and the remaining single subsample was used as the test dataset. The process was repeated until all groups have served as the validation data set. The results can be used for comparisons between the predicted and observed values. As validation measurements in this step of tuning QRF, we used the root mean square error (RMSE).

To develop the spatial version of QRF, residuals from the fitted QRF model were interpolated to prediction grids using ordinary kriging (Oliver and Webster, 2014), and the interpolated residuals were added to the QRF prediction results for obtaining the spatial QRF (sQRF) prediction. The spatial version of QRF is expressed as follows,

$$P_{sQRF}(s_0) = P_{QRF}(s_0) + \varepsilon(s_0)$$

where $P_{sQRF}(s_0)$ is the predicted value at location s_0, $P_{sQRF}(s_0)$ is the trend fitted by regular QRF, and $\varepsilon(s_0)$ are the kriged residuals by ordinary kriging.

To assess the model performance, the results of sQRF with spatial RF (sRF), RK and LR were compared to those of the case study using the same explanatory variables for all models. The analysis was implemented using the R packages "stats" for LR, "gstat" for RK and "caret" for RF. In sRF and RK, the residuals of an RF and regression liner model, respectively, were kriged and added to the predicted values in the same way as in sQRF. The "gstat" package allows the adjustment of semivariograms and kriging interpolation. The prediction error, which is the difference between the observed value for the dependent variable at a given location and the predicted value at the same location, was calculated for all compared models.

2.4. Model validation

The performance of the models developed using the selection methods mentioned above was evaluated using a 10-fold cross-validation. Mean absolute error (MAE), root mean squared error (RMSE), mean average percentage error (MAPE), coefficient of determination R^2, and amount of variance explained (AVE) were used to assess the performance of the tested methods. The following equations

were used to calculate these quality measurements:

$$R^2 = 1 - \frac{\sum_1^n \left(y_i - \hat{y}_i\right)^2}{\sum_1^n (y_i - \bar{y})^2}$$

where y_i is the actual land value, and $\hat{y}_i$ is the predicted value for i th property.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{\hat{y}}_i\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\hat{y}}_i\right)^2}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

$$AVE = 1 - \frac{\sum_1^n \left(y_i - \hat{y}_i\right)^2}{\sum_1^n \left(y_i - \underline{\hat{y}}\right)^2}$$

where $\underline{\hat{y}}$ is the mean of the target variable (land value) estimated from the validation sample.

Additionally, a graphical comparison was conducted using Taylor diagrams (Taylor, 2001), which enables simultaneous visualization of several metrics of model performance using correlation coefficients, RMSE, and standard deviations (SD) (Choubin et al., 2017).

The measures of uncertainty of sQRF and RK were evaluated and compared using the prediction interval coverage probability (PICP) (Shrestha and Solomatine, 2006). PICP is the probability that the observed values are within the prediction intervals (which is computed for a prediction level of 1– α (e.g., 95%). The method with a PICP near the confidence level (i.e., 95% with some tolerance) is the best method.

$$PICP = \frac{1}{n}\sum_{t=1}^{n} C, C = \left\{\begin{array}{l} 1, PL_t^{upper} < y_t < PL_t^{lower} \\ 0, otherwise \end{array}\right\}$$

where y_t is observed value, PL_t^{lower} and PL_t^{upper} are lower and upper prediction limits, respectively. Then for a given α, the PICP was computed and plotted against the confidence level to obtain the "accuracy plot" (Goovaerts, 2001) for which a 1:1 relationship is expected.

Another way to assess the quality of a model applied to spatial data is by calculating the degree of residual spatial autocorrelation (RSA) (Georganos et al., 2019). To investigate spatial correlation we employed the commonly used Moran's I Index (MI) (Moran, 1948) to assess the level of residual autocorrelation in incrementally increasing distance ranges.

$$MI = \frac{n\sum_i\sum_j w_{ij} z_i z_j}{M\sum_{i=1}^{n} z_i^2}$$

where n is the number of data points, $z_i = y_i - \bar{y}$; $\bar{y}$ is the mean value of y, $M = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$ and w_{ij} is the degree of spatial association between the points i and j (Kalogirou and Hatzichristos, 2007). The MI value ranges between −1 and 1, with values higher than 0 implying a positive spatial autocorrelation.

3. Results

3.1. Land unit values

The sampled LUVs (USD ha^{-1}) ranged from 1 to 25,946 USD ha^{-1}, with a high coefficient of variation (72%). The minimum value corresponded to areas with surface water, flooded or saline areas, which are considered as having no market value for valuation purposes; however, these are important points for the fit of the model and for obtaining a map of values along a continuum. The mean and median were similar, 6549 and 6570 USD ha^{-1}, respectively. Twenty-five percent of the samples had a value less than or equal to 2971 USD ha^{-1}, whereas another 25% had values greater than or equal to 9000 USD ha^{-1}.

3.2. Fitting mass appraisal models and LUV prediction

The degrees of RSA, calculated by MI at incrementing spatial scales for LR, RK, sRF and sQRF are shown in Fig. 2. The maximum MI value was reached at 5000 m in all models, except in RK (1000 m). The residuals from LR showed the highest autocorrelation for the entire range of distances evaluated; sQRF presented the lowest MI values.

Model performances of the fitted regression models did not vary widely in terms of global prediction error (Table 2). However, the model estimated with the sQRF algorithm produced better results than the classical LR model, particularly in terms of LUV predictions in unsampled properties (smallest RMSE).

Mean absolute error (MAE), mean average percentage error (MAPE), root mean squared error (RMSE), coefficient of determination (R^2) and amount of variance explained (AVE).

The visualization of the model performance using the Taylor diagram is presented in Fig. 3, where differences are represented by the standard deviation (SD) of the observed and predicted values. The plot shows that the standard deviation of the observed values is close to 4700 (USD ha^{-1}) (marked as 'observed' on the x-axis). The magnitude of the variability was measured as the radial distance from the origin of the plot. Thus, the predicted values from all models had lower variability than the observed values. The dot representing the LR model is close to the dashed line; therefore, the values predicted with this method show the closest variability to that of the observations. The correlation coefficient is shown on the arc, with dots (models) that lie closest to the x-axis being those that have the highest correlation. The best performing model, given the highest correlation coefficients, was sQRF (more than 0.85). The concentric dashed line originating from the 'observed' mark shows the value of the centered RMSE, showing that sQRF has the lowest prediction error (close to 2500 USD ha^{-1}), whereas the LR model has an RMSE of about 3100 USD ha^{-1}.

3.3. Mapping LUVs

The sQRF model was used for province-wide prediction of LUVs and the resulting spatial distribution was visualized as a map of LUVs

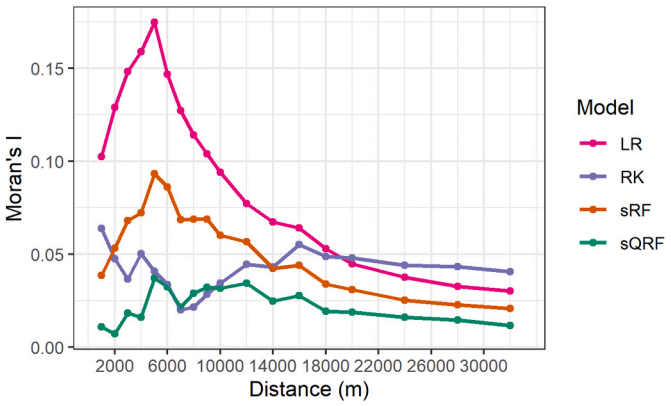

Fig. 2. Moran's I Index at increasing spatial distances of four regression models: Linear Regression (LR), Regression Kriging (RK), Spatial Random Forest (sRF), Spatial Quantile Regression Forest (sQRF).

Table 2
Global performance of mass appraisal models of farmlands in Córdoba, Argentina.

Model	MAE ($US. ha^{-1})	R^2 (%)	AVE (%)	MAPE (%)	RMSE ($US.ha^{-1})
Linear Regression (LR)	1819	57	53	34	3177
Regression Kriging (RK)	1726	61	60	34	2941
Spatial Random Forest (sRF)	1490	72	70	28	2534
Spatial Quantile Regression Forest (sQRF)	1413	73	73	25	2417

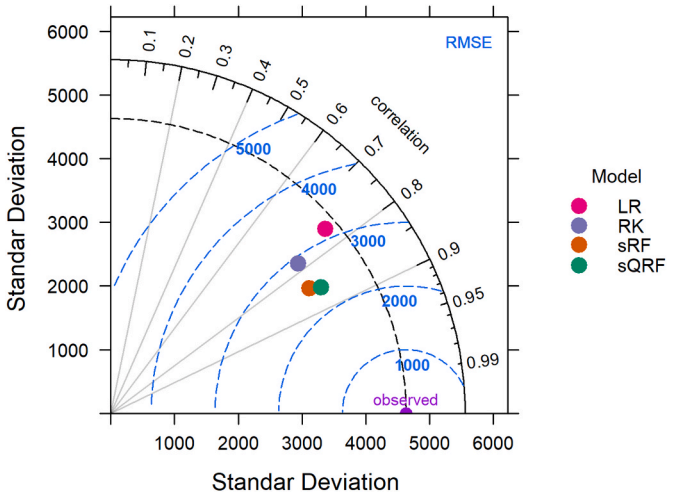

Fig. 3. Comparison of the performances of the mass appraisal models using a Taylor diagram. Linear Regression (LR), Regression Kriging (RK), Spatial Random Forest (sRF), Spatial Quantile Regression Forest (sQRF).

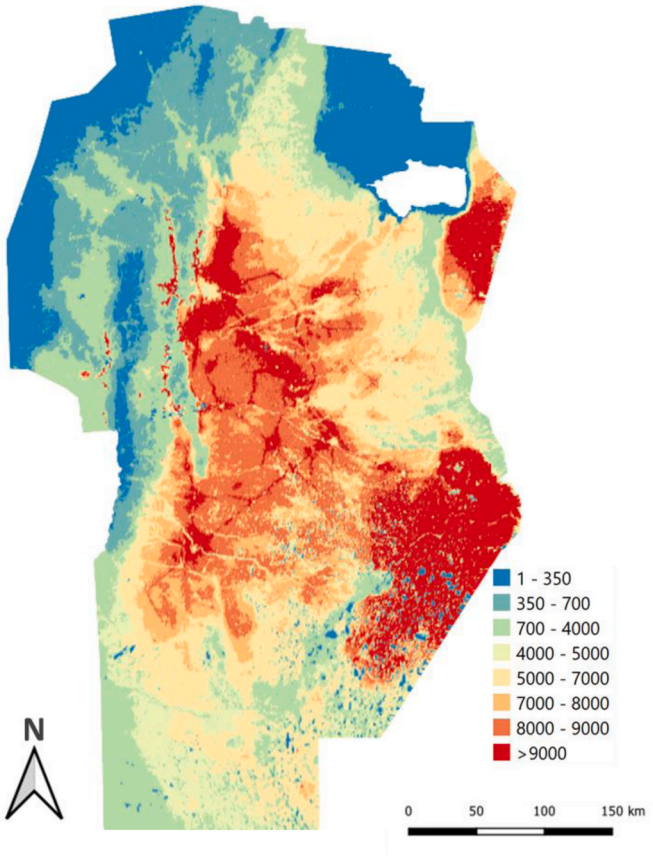

Fig. 4. Spatial variability of land unit values (USD ha^{-1}) of Córdoba, Argentina.

(Fig. 4). The map shows high LUV variability and a clear spatially structured pattern. Values increased from west to east, with the highest values corresponding to the southwest of the province. The values of the core agricultural area (southwest and east) amounted to approximately 9500 USD ha^{-1}, whereas in the central agricultural area, values were 7800 USD ha^{-1}. In the sierras area (west), characterized by livestock production and fruit and horticultural crops, LUV was about 2200 USD ha^{-1}. In the same area, under a marginal production, values were approximately 500 USD ha^{-1}.

The standard deviation for the prediction at a given site i, $\sigma(y_i)$ was computed from the estimated quantile associated with the lower and upper limits of a 68.27% prediction interval (Hengl et al., 2018).

$$\sigma sQRF(y_i) \approx \frac{\widehat{y}_{q=0.841(y_i)} - \widehat{y}_{q=0.159(y_i)}}{2}$$

The $\sigma(y_i)$ values were expressed as percentage of the predicted LUVs and were mapped in space (Fig. 5). The maps of prediction uncertainty exhibited spatial patterns closely related to the density of measured LUVs. The higher prediction errors (in red in the figure) were found mainly in the western and northern areas, where fewer observations were available. Furthermore, in areas of high uncertainty, the LUVs are expected to vary over short distances.

The results of the PICP analysis indicate that, at the desired confidence level of 95%, 89 and 90% of all observations fitted within the expected prediction interval for sQRF and RK, respectively. Therefore, the empirical uncertainty model is optimal for both predictive models, RK and sQRF. Furthermore, with each successive decrease in the confidence levels, a near corresponding decrease in the PICP is observed for both models, indicating a required outcome in terms of sensitivity of the PI to changing confidence levels (Fig. 6).

3.4. Variable importance

The level of variable importance measures according to sQRF is shown in Fig. 7. The higher the percentage increase in node purity during the modeling process of a variable, the higher the importance of that feature in the prediction model. The five most important variables for predicting LUVs were soil productivity index (SPI), zones of appraisal value (ZAV), edaphoclimatic area, annual crops area, and lease value. All variables, except Palmer Drought severity index, were spatially structured, as suggested by a significant Moran's Index ($P < 0.0001$).

4. Discussion

The need for transparent, realistic, and updated land valuation is essential in all aspects of land administration, especially when numerous land parcels are involved (Bencure et al., 2019). Mass appraisal for the determination of land value has multiple applications, including territorial taxation. There are also many other actions in rural properties that also need correct determination of land value, such as: financing, expropriations, indemnities (in case of creation of conservation units or even in environmental disasters), buying and selling of real estate, and land reform (Uberti 2018). Therefore, reliable and accurate rural land values are critical to the development of land policies. Our goal is to introduce a computing intensive methodology for assessing rural land values from neighboring data of the rural lands as an innovative land valuation model.

Several quantitative valuation methods for mass appraisal that are useful in metropolitan areas are not as useful in non-urban areas. Particularly, rural LUVs exhibit high variability across a territory and

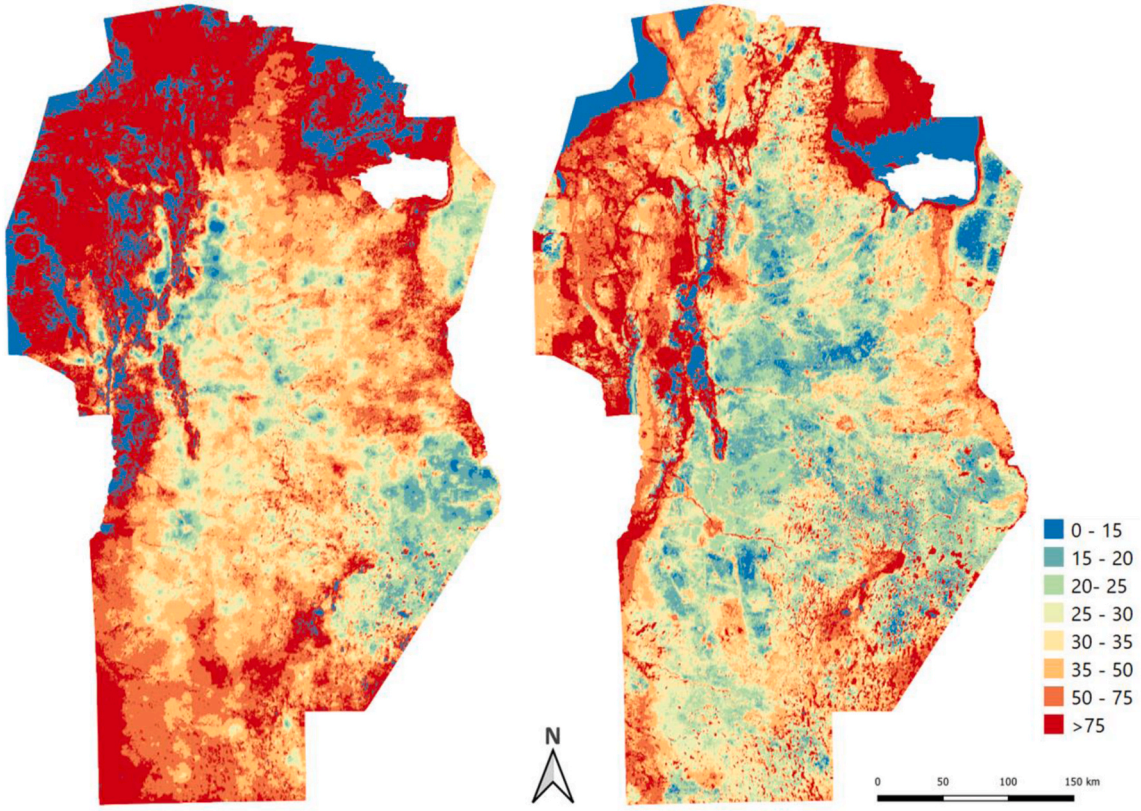

Fig. 5. Relative standard deviation of land unit values (LUVs) (expressed as % of predicted value) in Córdoba, Argentina. Predictions of LUVs were obtained by regression kriging (left) and spatial quantile regression forest (right).

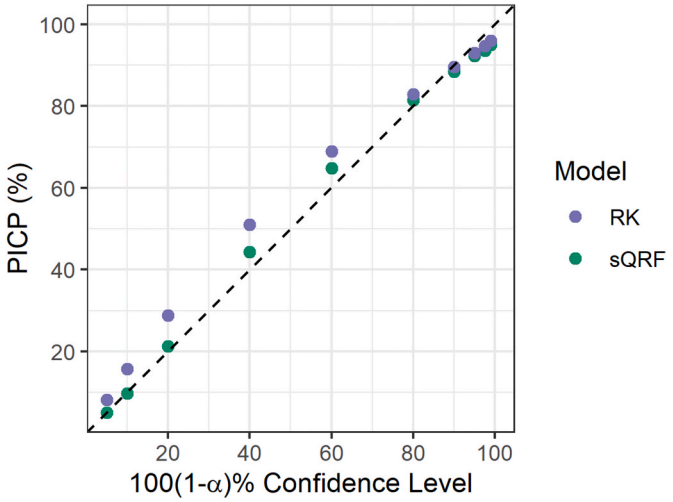

Fig. 6. Prediction interval coverage probability plots (PICPs) for land unit values.

usually a strong spatial behavior, mainly associated with the natural resources and variability in local economies. The abundant amount of ancillary information from the neighboring area of the land to be priced can be used to improve mass appraisal. Automated or easy to compute machine learning models combined with geographical ancillary information becomes an interesting approach to the determination of consistent LUVs of rural lands.

Some machine learning methods, including tree-based regression models, have been used for rural mass appraisal, but without dealing with the spatial covariability underlying the land value distributions. In this study, the performances of a version of QRF algorithm, a tree-based machine learning model, is modified to account for spatiality; the new algorithm was named sQRF. The paper innovation lies in the development of a spatially based machine learning model for mass appraisal of rural land. The algorithm is tested against other popular algorithms to build predictive models, such as classical random forest, and the multiple regression model with and without spatial constraints. Our findings suggest that sQRF performed well in delivering a digital map of rural land prices using commonly available explanatory variables. Moreover, sQRF provided information about the relative importance of ancillary variables and uncertainty measurements to better interpret the predicted LUVs. These attributes evidence the important role that sQRF algorithm may play in valuation of rural land.

In our study, we assessed the goodness of fit of sQRF and of alternative regression models for mass appraisal of farmland values trough global and pointwise prediction error calculation. Differences in performance were observed between LR and RK. From a theoretical point of view, a perfect regression model should have an R^2 close to 1 to fit the measurements, and it is used to predict new values, i.e. unsampled ones, a null prediction error. However, there is no such thing as a perfect model to be used in digital mapping of spatial variability (Guevara and Olmedo, 2018). Depending on the situation, some performance measures might be more appropriate than others. Hence, there is not a single measure to select a model, with model evaluation requiring a combination of performance measures (Chang and Hanna, 2004). When the LR was used to model land values, the residuals exhibited a statistically significant spatial structure, suggesting the need to incorporate the underlying spatial autocorrelation in the modeling, whereas for RK, a regression model for spatial data, the magnitude of the autocorrelation in the residuals decreased significantly (Fig. 2). The spatial autocorrelation, especially if still existent in the residuals, indicates that the predictions may be biased, which is suboptimal to obtain spatial

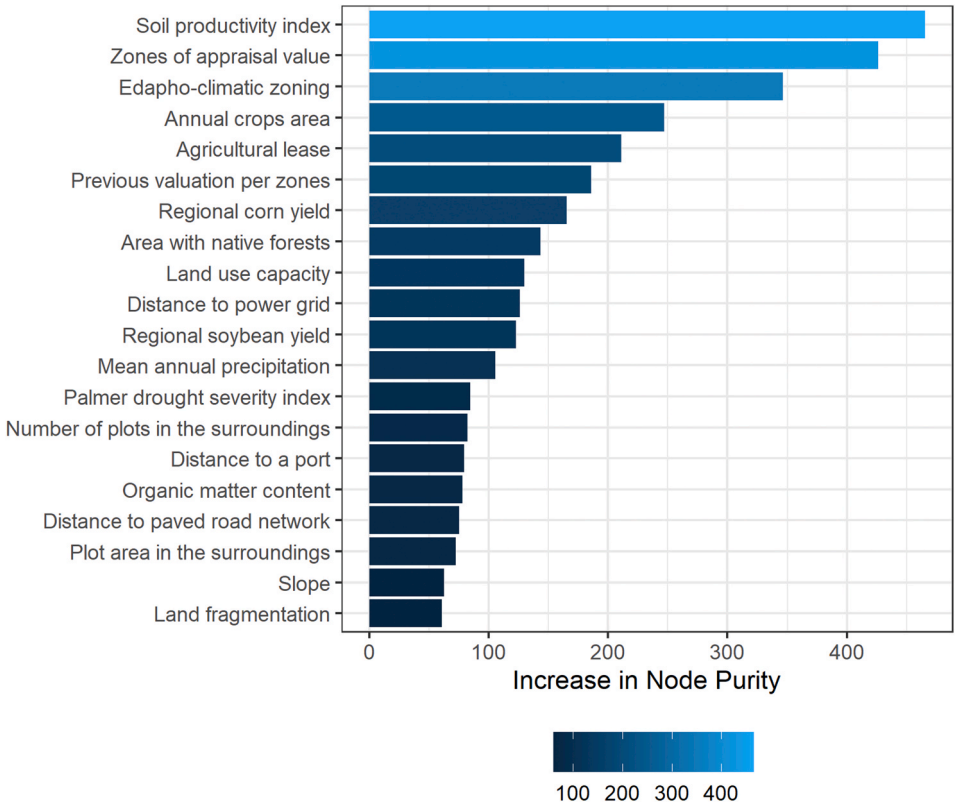

Fig. 7. Importance plot of top 20 predictive variables based on the spatial quantile regression forest.

predictions (Hengl et al., 2018). High correlation in the residuals is a typical phenomenon when a model has not been specified correctly (Georganos et al., 2019), usually because important explanatory variables with high spatial variability are missing or because they fail to account for spatial dependency in the observed data. Several site covariates included in the regression model, such as distance to power grid or distance to the port, can operate as spatial predictors.

The addition of kriging residuals to the model predictions improved the predictive accuracy of the fitted models (Table 1). The spatial autocorrelation in the residuals from sQRF, obtained by calculating MI at incrementing spatial scales, was low (MI values < 0.04) for the entire range of distances evaluated (Fig. 2). Modeling with sQRF methodology allowed us to capture complex relationships between explanatory variables defining the land value and improved prediction errors by 24% with respect to the classical multiple regression model.

The sQRF showed the highest AVE and the smallest RMSE, meaning that it was the best method for adjusting a full model (model with all data), but also a predictive model (model evaluated with data not used in calibration). Both measurements are calculated from the difference between observed and predicted values and are then averaged to obtain global accuracy measures. The fraction of the variation in the data that is explained by the fitted model is a measure of model efficiency to capture the response variable variability; that fraction was highest with sQRF, as observed by the AVE values and R^2 assessing the agreement between the model-based values and the measured values.

Since sQRF infers the full conditional distribution of a response variable, it was used to build uncertainty measurements to complement the interpretation of LUV maps. The sQRF method produced a prediction interval from the quantiles of the conditional distribution. The uncertainty variance of the sQRF model was lower than that of RK; therefore, more reliable predictions are expected from the spatial version of QRF.

One advantage of machine learning models based on random trees is that they enable assessment of the importance of the explanatory variables (Rahmati et al., 2019). In the case study, the variables that most contributed to determine LUVs were SPI and ZAV (Fig. 7). The explanatory variable SPI was developed by FAO (Riquier et al., 1970) and then adapted to the conditions of the study region. The variable ZAV corresponds to reference values assigned by expert real estate appraisers. Thus, expert knowledge is crucial for making good LUV predictions. The sQRF algorithm revealed that different types of factors controlling the variability of LUVs include cadastral (plot area and plot number in surrounding, previous valuation per zones, territorial fragmentation), soil (SPI, edaphoclimatic zones, land use capacity, organic matter content), vegetation or land cover (annual crop area, area with native forest), topography (slope), climate (rainfall, drought index), location (distance to main road, distance to power grid, distance to the port), and economic factors (lease value, corn yield, soybean yield). The influence of these variables on the LUVs has also been recognized in previous studies (Pienaar, 2013; van der Walt and Boshoff, 2017). Using spatial LR models, Uberti et al. (2018) found that variables similar to SPI were significant at 0.1 (p-value) for mass appraisal of farmland.

A strong spatial correlation structure and high variability of LUVs was observed in Córdoba. The highest values in the southeastern sector of the province were associated with the extensive agriculture core production area (soybean, corn, wheat) located in the humid Pampas of Argentina, characterized by the best soils. Another large-scale spatial trend was observed with the increase of LUVs to the northeastern region of the province, an area of extensive agriculture and milk production. Moreover, high values associated with road and urban infrastructure (proximity to localities/central economies) were found due to the high prediction resolution (prediction grids of 500 × 500 m). The less developed areas are located to the northwest of the province, where on average, LUVs were lower. Patches, i.e., small areas with contrasting values with respect to the neighboring areas, were observed both in the southeast and in the west. In the southeast, those areas are associated with frequent floods and therefore have a low market value. In the west, a marginal area for agricultural production, patches of highest LUVs corresponded to corridors of high value for tourism.

The LUVs produced by sQRF correspond to the re-valuation conducted in 2019, which is effective for 2020. The preceding valuation, performed in 1994, was conducted following a specific territory structure, in agreement with land production characteristics and was not digitally mapped. As a result, the value per hectare was assigned to only 833 valued zones (polygons). The implementation of sQRF for predicting LUVs allows us to improve the spatial resolution and is a particular alternative to ensure the sustainability of a correct and updated land valuation that incorporates the rapid transformations occurring in the territory (Piumetto et al., 2019). In addition, the new methodology helps to reduce costs and time involved in the land revaluation process. According to the uncertainty map, sQRF produced good predictions with more than half of the predicted LUVs, with prediction errors below 30% of the observed LUVs. However, a spatial pattern was also observed in the standard deviation of the predicted LUVs. There is a high heterogeneity of LUVs in the west, where besides tourist hot spots, there are both lands destined for marginal agriculture and areas under irrigation with intensive agriculture. Although the southeastern and eastern areas are more homogeneous, uncertainty of LUVs was found in some areas probably associated with floods. Therefore, the land values are drastically reduced in comparison to similar non-flooded sites, generating high variability in short distances. The results showed that with a sample of 2% of all land polygons in Cordoba, sQRF provided an adequate model tool for appraising the LUVs of all the area, generating LUV digital maps and additional uncertainty maps for the LUV predictions. The advantages of having an updated valuation of these lands include: i) greater transparency and legal certainty for owners; ii) fair taxation; iii) more information to design public policies and decide on production investments; iv) better tools to implement tax relief for those affected by natural disasters, and v) the possibility of measuring –social and economic— impacts of climate change. Land is the most precious and limited non-renewable resource; yet, it is one of the most exploited and undervalued natural resources (Bencure et al., 2019). Analysis of the land market can be extremely useful for decisionmakers, since it can be used for designing measures aimed at preserving several types of resources spread across the rural territory (Sardaro et al., 2020). From LUVs obtained by applying SQRF in Córdoba, the valuation of more than two million properties in the province were updated. The values of rural land corresponding to 2019 increased on average eight times with respect to the previous valuation. The results were approved by a resolution of the Cadastral Agency of the government of Córdoba, and 70% of the values obtained from this study were adopted for cadastral valuation. Thus, valuations were updated according to the market levels, and value determination was separated from tax policy. The new tax was applied gradually, considering that the valuations were outdated. Regarding tax management, behavior of taxpayers was not affected, and claims were below 0.05%. A correct rural land valuation as the source for estimation of real estate tax contributes to a reduction in tax distortion and in economic losses.

5. Conclusions

In this paper, we model the relationship between a set of environmental variables and rural land values. To this end, we develop a spatial version of a machine learning algorithm, as a new method for LUV predictions. The main advantage of applying the sQRF methodology for mass appraisal is the possibility to handle big data and its high predictive accuracy. The sQRF outperformed LR, RK, and sRF in delivering LUV predictions, as shown by several validation measurements. The procedure was able to derive better predictions of LUVs at unsampled sites than the classical or even the spatial linear regression model. The algorithm sQRF uses site-specific explanatory variables and LUVs from neighboring sites to improve LUV predictions. Moreover, it has the ability to yield an uncertainty measure for the predicted LUVs. The findings showed that soil quality, was the most important environmental variable in explaining farmland values in Córdoba. The methodology can be used for objective assessment of LUVs in other territories. The sQRF method can be implemented using free software and widely available territorial data; however, as other supervised machine learning methods, it lacks automation. Future research on automatic optimization of sQRF parameters will facilitate its use in digital economy. This study acknowledges the importance of a correct assessment of the LUVs.

Author statement

Mariano Córdoba: Writing - Original Draft, Investigation, Formal analysis, Methodology, Validation, Visualization, Software. **Juan Pablo Carranza:** Writing - Review & Editing. **Mario Piumetto:** Writing - Review & Editing, Project administration, Funding acquisition. **Federico Monzani:** Writing - Review & Editing. **Mónica Balzarini**: Writing - Review & Editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The information used in the present work was generated within the frame of the Estudio Territorial Inmobiliario (ETI) of the Province of Córdoba, Argentina, funded jointly by the United Nations Development Programme (PNUD AR/16/005) and the Government of the province of Córdoba. The project was coordinated by the Secretaria de Ingresos Públicos and the Dirección General de Catastro, both agencies depending on the Ministerio de Finanzas. We are grateful to all the ETI team members, for their participation in the different stages of the project. The statistical research was partially funded by Consejo Nacional de Investigaciones Cientificas y Técnicas (CONICET).

References

Antipov, E.A., Pokryshevskaya, E.B., 2012. Expert Systems with Applications Mass appraisal of residential apartments : an application of Random forest for valuation and a CART-based approach for model diagnostics. Expert Syst. Appl. 39, 1772–1778. https://doi.org/10.1016/j.eswa.2011.08.077.

Bencure, J.C., Tripathi, N.K., Miyazaki, H., Ninsawat, S., Kim, S.M., 2019. Development of an innovative land valuation model (iLVM) for mass appraisal application in suburban areas using AHP: an integration of theoretical and practical approaches. Sustain. Times 11. https://doi.org/10.3390/su11133731.

Bidanset, P.E., Lombard, J.R., 2014. Evaluating spatial model accuracy in mass real estate appraisal A comparison of geographically weighted regression and the spatial lag model. Source: Cityscape 16, 169–182.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. CRC press.

Caballer, V., 2008. Valoración Agraria. Teoría Y Práctica, fifth ed. Mundi-Prensa Libros.

Čeh, M., Kilibarda, M., Lisec, A., Bajat, B., 2018. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. ISPRS Int. J. Geo-Inf. 7, 168. https://doi.org/10.3390/ijgi7050168.

Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87 https://doi.org/10.1007/s00703-003-0070-7.

Choubin, B., Malekian, A., Samadi, S., Khalighi-Sigaroodi, S., Sajedi-Hosseini, F., 2017. An ensemble forecast of semi-arid rainfall using large-scale climate predictors. Meteorol. Appl. 24, 376–386. https://doi.org/10.1002/met.1635.

Choumert, J., Phelinas, P., Choumert, J., Phelinas, P., 2014. Determinants of Agricultural Land Values in Argentina to Cite This Version : HAL Id : Halshs-01027502 Determinants of Agricultural Land Values in Argentina.

Demetriou, D., 2017. A spatially based artificial neural network mass valuation model for land consolidation. Environ. Plan. B Urban Anal. City Sci. 44, 864–883. https://doi.org/10.1177/0265813516652115.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 1–16. https://doi.org/10.1080/10106049.2019.1595177.

Georgiadis, A., 2018. Real estate valuation using regression models and artificial neural networks: an applied study in Thessaloniki. Rel. Int. J. Real Estate L. Plan. 1, 292–303.
Giannini Kurina, F., Hang, S., Cordoba, M.A., Negro, G.J., Balzarini, M.G., 2018. Enhancing edaphoclimatic zoning by adding multivariate spatial statistics to regional data. Geoderma 310, 170–177. https://doi.org/10.1016/j.geoderma.2017.09.011.
Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103, 3–26. https://doi.org/10.1016/S0016-7061(01)00067-2.
Guevara, M., Olmedo, G.F., 2018. Model evaluation in digital soil mapping. In: Yusuf Yigini, Guillermo Federico Olmedo, Stephanie Reiter, Rainer Baritz, KostiantynViatkin, RonaldVargas. GlobalSoilPartnership, F.O. Of the U.N., second ed. FAO, Rome, p. 220. Soil Organic Carbon Mapping Cookbook.
Guo, P.T., Li, M.F., Luo, W., Tang, Q.F., Liu, Z.W., Lin, Z.M., 2015. Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach. Geoderma 238, 49–59. https://doi.org/10.1016/j.geoderma.2014.08.009, 237.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of africa at 250 m resolution: random forests significantly improve current predictions. PloS One 10, e0125814. https://doi.org/10.1371/journal.pone.0125814.
Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33, 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001.
Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. https://doi.org/10.7717/peerj.5518.
IAAO, 2017. Standard on Mass Appraisal of Real Property (SMARP) (Kansas).
Jahanshiri, E., Buyong, T., Shariff, A.R.M., 2011. A review of property mass valuation models. Pertanika J. Sci. Technol. 19, 23–30.
Kalogirou, S., Hatzichristos, T., 2007. A spatial modelling framework for income estimation. Spatial Econ. Anal. 2, 297–316. https://doi.org/10.1080/17421770701576921.
Keys, R., 1981. Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. 29, 1153–1160. https://doi.org/10.1109/TASSP.1981.1163711.
Kontrimas, V., Verikas, A., 2011. The mass appraisal of the real estate by computational intelligence. Appl. Soft Comput. 11, 443–448. https://doi.org/10.1016/j.asoc.2009.12.003.
Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the boruta package. J. Stat. Software 36, 1–13. https://doi.org/10.18637/jss.v036.i11.
McCluskey, W.J., McCord, M., Davis, P.T., Haran, M., McIlhatton, D., 2013. Prediction accuracy in mass appraisal: a comparison of modern approaches. J. Property Res. 30, 239–265. https://doi.org/10.1080/09599916.2013.781204.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.
Moran, P.A.P., 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B 10, 243–251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x.
Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: computing and modelling variograms and kriging. Catena 113, 56–69. https://doi.org/10.1016/j.catena.2013.09.006.
Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. Nature 540, 418–422. https://doi.org/10.1038/nature20584.
Pienaar, P., 2013. Farm valuations in practice. Agri Land Price Index.
Piumetto, M., 2020. La innovación como clave para la actualización de valores: el caso de la Provincia de Córdoba. In: Eguino, H., Erba, D. (Eds.), Catastro, Valoración Inmobiliaria Y Tributación Municipal: Experiencias Para Mejorar Su Articulación Y Efectividad. Banco Interamericano de Desarrollo, Washington. D.C., p. 64
Piumetto, M., Garcìa, G., Monayar, V., Carranza, J., Morales, H., Nasjleti, T., Menéndez, A., 2019. Modernización de la Valuación Masiva de la Tierra en la provincia de Córdoba a través de técnicas de aprendizaje computacional. Rev. la Fac. Ciencias Exactas, Físicas y Nat. 6, 49–52.
R Core Team, 2020. R: A Language and Environment for Statistical Computing.
Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Ahmad, B. Bin, Tien Bui, D., 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. Sci. Total Environ. 688, 855–866. https://doi.org/10.1016/j.scitotenv.2019.06.320.
Riquier, J., Bramao, D.L., Cornet, J.P., 1970. A New System of Soil Appraisal in Terms of Actual and Potential Productivity.
Sardaro, R., La Sala, P., Roselli, L., 2020. How does the land market capitalize environmental, historical and cultural components in rural areas? Evidences from Italy. J. Environ. Manag. 269, 110776. https://doi.org/10.1016/j.jenvman.2020.110776.
Sesli, F.A., 2015. Creating real estate maps by using GIS: a case study of Atakum-Samsun/Turkey. Acta Montan. Slovaca 20, 260–270. https://doi.org/10.3390/ams20040260.
Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. Neural Network. 19, 225–235. https://doi.org/10.1016/j.neunet.2006.01.012.
Simões, J.C.M., Ferreira, F.A.F., Peris-Ortiz, M., Ferreira, J.J.M., 2020. A cognition-driven framework for the evaluation of startups in the digital economy. Manag. Decis. 58, 2327–2347. https://doi.org/10.1108/MD-09-2019-1253.
Szatmári, G., Pásztor, L., 2019. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337, 1329–1340. https://doi.org/10.1016/j.geoderma.2018.09.008.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.
Uberti, M.S., Antonio, M., Antunes, H., Debiasi, P., Tassinari, W., 2018. Land Use Policy Mass appraisal of farmland using classical econometrics and spatial modeling. Land Use Pol. 72, 161–170. https://doi.org/10.1016/j.landusepol.2017.12.044.
van der Walt, K., Boshoff, D., 2017. An analysis of the use of mass appraisal methods for agricultural properties. Acta Structilia 24, 44–76. https://doi.org/10.18820/24150487/as24i2.2.
Wang, D., Li, V.J., 2019. Mass appraisal models of real estate in the 21st century: a systematic literature review. Sustain. Times 11, 1–14. https://doi.org/10.3390/su11247006.
Yacim, J.A., Boshoff, D.G.B., 2018. Impact of artificial neural networks training algorithms on accurate prediction of property values. J. R. Estate Res. 40, 375–418. https://doi.org/10.5555/0896-5803.40.3.375.
Zhang, R., Du, Q., Geng, J., Liu, B., Huang, Y., 2015. An improved spatial error model for the mass appraisal of commercial real estate based on spatial analysis : shenzhen as a case study. Habitat Int. 46, 196–205. https://doi.org/10.1016/j.habitatint.2014.12.001.