

Article

High-Resolution Mapping and Assessment of Salt-Affectedness on Arable Lands by the Combination of Ensemble Learning and Multivariate Geostatistics

Fatemeh Hateffard ¹ , Kitti Balog ^{2,*} , Tibor Tóth ² , János Mészáros ² , Mátyás Árvai ² ,
Zsófia Adrienn Kovács ², Nóra Szűcs-Vásárhelyi ² , Sándor Koós ², Péter László ², Tibor József Novák ¹ ,
László Pásztor ²  and Gábor Szatmári ^{2,3} 

¹ Department of Landscape Protection and Environmental Geography, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

² Institute for Soil Sciences, Centre for Agricultural Research, Herman Ottó út 15, H-1022 Budapest, Hungary

³ Department of Physical Geography and Geoinformatics, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

* Correspondence: balog.kitti@atk.hu

Abstract: Soil salinization is one of the main threats to soils worldwide, which has serious impacts on soil functions. Our objective was to map and assess salt-affectedness on arable land (0.85 km²) in Hungary, with high spatial resolution, using a combination of ensemble machine learning and multivariate geostatistics on three salt-affected soil indicators (i.e., alkalinity, electrical conductivity, and sodium adsorption ratio ($n = 85$ soil samples)). Ensemble modelling with five base learners (i.e., random forest, extreme gradient boosting, support vector machine, neural network, and generalized linear model) was carried out and the results showed that ensemble modelling outperformed the base learners for alkalinity and sodium adsorption ratio with R^2 values of 0.43 and 0.96, respectively, while only the random forest prediction was acceptable for electrical conductivity. Multivariate geostatistics was conducted on the stochastic residuals derived from machine learning modelling, as we could reasonably assume that there is spatial interdependence between the selected salt-affected soil indicators. We used 10-fold cross-validation to check the performance of the spatial predictions and uncertainty quantifications, which provided acceptable results for each selected salt-affected soil indicator (for pH value, electrical conductivity, and sodium adsorption ratio, the root mean square error values were 0.11, 0.86, and 0.22, respectively). Our results showed that the methodology applied in this study is efficient in mapping and assessing salt-affectedness on arable lands with high spatial resolution. A probability map for sodium adsorption ratio represents sodic soils exceeding a threshold value of 13, where they are more likely to have soil structure deterioration and water infiltration problems. This map can help the land user to select the appropriate agrotechnical operation for improving soil quality and yield.

Keywords: salt-affected soils; digital soil mapping; ensemble modelling; geostatistics; uncertainty assessment; satellite remote sensing; unpiloted aerial vehicle



Citation: Hateffard, F.; Balog, K.; Tóth, T.; Mészáros, J.; Árvai, M.; Kovács, Z.A.; Szűcs-Vásárhelyi, N.; Koós, S.; László, P.; Novák, T.J.; et al. High-Resolution Mapping and Assessment of Salt-Affectedness on Arable Lands by the Combination of Ensemble Learning and Multivariate Geostatistics. *Agronomy* **2022**, *12*, 1858. <https://doi.org/10.3390/agronomy12081858>

Academic Editor: Karsten Schmidt

Received: 23 June 2022

Accepted: 4 August 2022

Published: 6 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Among the major challenges that humanity faces, food security is the most important, and due to several unfavorable tendencies, such as climate change, population pressure, and belligerence, long-term agricultural sustainability is very much threatened in our times. Soil, as the basis of agronomy, provides several services and most 'soil ecosystem services' are also threatened by those tendencies; therefore, efficiency of agricultural operations, such as tillage, fertilizing, and irrigation, must be improved to meet the new challenges.

Soil salinization is considered as one of the main threats to soils, not just in Europe but around the world as well [1,2]. It is well known that surplus salt in soils has serious

impacts on ecosystem services provided by soils, which leads to a series of consequences, such as a decrease in agricultural productivity [3], reduced soil fertility [4], accelerating soil erosion [5], degradation of soil structure [6], decreasing ability to act as a buffer, and filter against pollutants [7], just to mention a few. In the conditions of the Hungarian Plain, soil salinization/sodification/alkalinization are widespread degradation processes from an agricultural point of view. However, we should note that areas with salt-affected soils (SAS) can also serve as unique and “ex lege” protected habitats (e.g., grasslands, hayfields, marshes, reed-lands, lakes) for animals (mainly birds) and plants.

In 2019, the Global Soil Partnership of the Food and Agricultural Organization launched the Global Map of Salt-affected Soils (GSAS map) international initiative, which was aimed at updating the global and country-level information on salt-affected soils and lay the ground for future monitoring [8]. This GSAS map was released in October 2021, with contributions from over 118 countries, including Hungary [9], and shows that more than 4,240,000 km² of topsoil (0–30 cm) and 8,330,000 km² of subsoil (30–100 cm) are salt-affected worldwide. Although the Global Map of Salt-affected Soils is an important product, which can be used for identifying salt-affected soils or launching monitoring programs, with a resolution of 1 km, it can give little or rough information at the country level, especially when spatial planning is targeted.

The production of large-scale maps is justified due to the high variability and site diversity in natural saline areas. Microtopography plays an important role in the diversity of soil properties in these areas, which is also indirectly reflected in the formation of vegetation pattern. Among soil-forming factors [10], elevation was found to be the main influential factor, closely correlated with biomass value (10-year average), its range, and salinity [11]. In local depressions, increased infiltration and capillary action (from shallow salty groundwater) can be seen, along with increased salinity and, therefore, reduced productivity. Due to the regularly occurring waterlogging during wet springs with shallow groundwater table, biomass and normalized difference vegetation index (NDVI) were generally low and spatially heterogeneous, while in the case of a dryer spring, biomass was larger and more homogeneous due to no waterlogging being observed and the groundwater table was found at a greater depth. Productivity proxy values were more stable and higher at higher elevation, avoiding being affected by yearly fluctuations in shallow groundwater table, waterlogging, and changes in the amount of precipitation. On the other hand, several former saline areas are used for agricultural purposes as a result of soil amelioration, but for these soils, productivity-sustaining management is essential. Precision farming can be a suitable method for this management, which requires knowledge of soil chemical parameters within the plot. This information can exclusively be ensured with large-scale maps. Furthermore, there is a need to map soil salinity/sodicity/alkalinity status in detail (100 × 100 m resolution) and collect new information on these for ensuring the subsidies linked to agricultural areas, with natural constraints from the European Union (EU) [12], due to the low limit for sodium that cannot be determined nor estimated on the basis of the formerly prepared maps. As a result, large areas may be excluded from statutory support.

In Hungary, there is a long tradition and history of studying salt-affected soils that is demonstrated by a huge number of monographs (e.g., [13–17]). Most of the areas with SAS can be found in the Great Hungarian Plain, an alluvial plain filled up with thick alluvial sediments on an ancient seabed. Later loess formation also took place here and the influence of shallow fluctuating saline-sodic groundwater, as well as permanent or temporary waterlogging created the conditions for SAS formation. Sodium ions, being considered as the most important factor, are either dissolved from Tertiary Era deposits into groundwater [18] or concentrated during consecutive drying and wetting of infiltrated water [19]. Systematic mapping of salt-affected soils has a history of more than a century in Hungary and, since then, a number of maps has been compiled at various scales (e.g., [14,20–22]).

Digital soil mapping (DSM) aims at providing spatial or spatio-temporal information on the soil for a wide range of disciplines [23,24], such as agronomy, rural development,

hydrology, and environmental sciences, just to name a few. Nowadays, advanced geostatistical and machine learning techniques, as well as their combinations, are in use for predicting and mapping the spatial or spatio-temporal distribution of various soil properties, functions, and services [25–31], which is frequently complemented by proper accuracy assessment [32] and spatial uncertainty prediction [33,34]. Very recently, ensemble machine learning, which shares a joint optimal predictive model obtained from a combination of two or more individual learners, is increasingly used in DSM to improve prediction performance (e.g., [35–38]). Ensemble modelling is able to provide robust and accurate predictions while reducing variance [39,40]; furthermore, ensembles successfully solve the issues that machine learning algorithms are dealing with, such as handling missing values, improving confidence estimation by weighing various variables, and considering the most important ones [41,42]. The amount of potentially available environmental covariates used in DSM is continuously increasing, mainly thanks to optical remote and proximal sensing [43]. Although digital elevation models and geomorphometric parameters (e.g., slope, topographic wetness index) derived from them are informative covariates in DSM, remote and proximal sensing can provide a huge amount of information on land surface, with a continuously increasing spatial, temporal, and spectral resolution [43–45]. Additionally, a combination of certain bands of multi- or hyperspectral images, such as thematic indices (e.g., normalized difference vegetation index, salinity index), can be obtained, which provides specific information for the given problem [9,46,47].

The objectives of this study were twofold. First, jointly modelling and mapping of selected SAS indicators, namely alkalinity, electrical conductivity (EC), and sodium adsorption ratio (SAR), with high spatial resolution, was targeted at an area of arable land, situated in Hungary, using ensemble machine learning and multivariate geostatistical techniques. We used multivariate geostatistics combined with ensemble machine learning, as our hypothesis was that the selected SAS indicators show spatial interdependency (i.e., they are cross-correlated in space) and, therefore, it is better to jointly model and map their spatial distribution. Second, based on the resulting maps, the assessment of alkalinity, salinity, and sodicity was targeted at the field scale, which could be used for spatial decision support, e.g., in precision agricultural operations.

2. Materials and Methods

2.1. Study Site

The study plot was selected by overlapping the maps of EU-supported arable lands and salt-affected soils in Hungary. The one having the largest contiguous area (0.85 km²) was selected for examination of all overlapping plots. This formerly saline/sodic, currently agriculturally used, plot is represented in Figure 1A. It is located in the former floodplain of the River Danube, outskirts of Dunavecse, Central Hungary. The plot has a ‘quasi’ rectangular shape with corner coordinates of 46°55′16″ N, 19°01′37″ E, 46° 55′17″ N, 19°02′12″ E, 46°55′55″ N, 19°01′41″ E, 46°55′49″ N, 19°02′12″ E and belongs to the microregion “Soltisik” according to Dövényi [48]. In this microregion, the annual precipitation average ranges from 530 to 550 mm with high temporal and spatial variability. Rainfall events alternate with longer dry periods, especially in spring and summer favoring the movement of salts. Total annual solar radiation is between 2000 and 2020 h, whereas long-term mean annual temperature is 10.4–10.5 °C [48]. The area has a negative water balance. Investigated plot is currently plowed, fertilized, and crop rotated (i.e., maize, sunflower, and barley) with relatively good yields (e.g., 10 tons per hectare of maize, county average: 7.27 tons per hectare). This plot was best suited for our study, as it has been managed consistently over the past 50 years, thereby facilitating the collection of data regarding the management and remote sensing data.

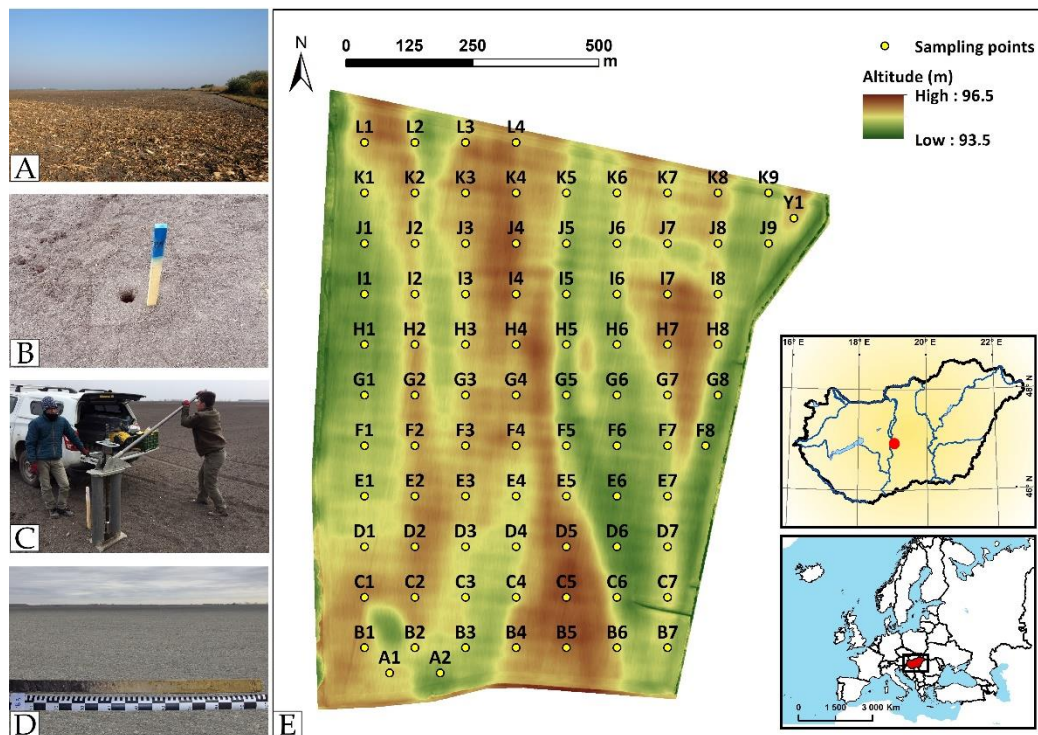


Figure 1. (A). Photo of the sampling plot (B). Photo of the field marking of the borehole (C). Photo of soil sampling with motorized hand drill (D). Photo of the undisturbed soil column lying in 10 cm diameter soil tube (E). Location of the study site in Central Hungary and its high-resolution digital elevation model coming from remote sensing (unpiloted aerial vehicle survey). Sampling points ($n = 85$).

2.2. Field Survey and Laboratory Analysis

In total, 85 soil tubes (10 cm in diameter plastic tubes, containing 1 m length undisturbed soil column, Figure 1D) from surface to 1 m depth were deepened inside the 0.85 km² studied area with motorized hand drill (Figure 1C). Sampling scheme followed a regular grid with sampling soil tubes from every 100 m (Figure 1E). Further, 10 locations out of the 85 grid points were deepened down to the groundwater table. Collected soil tubes were cut vertically at their centers and dissected (Figure 1D). Soil profiles were described according to IUSS Working Group, World Reference Base for Soil Resources [49]. Samples were taken from each genetic horizon for further analysis.

The study site was surveyed by an unpiloted aerial vehicle (UAV) carrying a visible-range Fujifilm camera (main features: APS-C sensor, 24Mp, focal length of 14mm f/2.8, AUTO ISO, automated exposure) to generate digital elevation model (DEM) with structure from motion technique at 10 cm spatial resolution. We performed the aerial surveys in a fully automatic flight mode above the study area with image over- and sidelap of 75%. Altitude of 140 m was found to be the most suitable for flight time and resolution. Temporary ground control targets were placed and their 3D coordinates were surveyed to make possible the orthorectification of raw images during image processing in Agisoft Metashape (<http://www.agisoft.com/downloads/installer/>, accessed on 1 June 2022). The final DEM was exported in 2 m resolution (Figure 1E) to be uniform with the other environmental covariates.

The selected SAS indicators, namely alkalinity, electrical conductivity (EC), and sodium adsorption ratio (SAR) were measured in the laboratory on the collected soil samples. Soil EC and alkalinity were measured as 1:2.5 soil:distilled water suspension with WTW Multi 350i combined electrode. Na concentration (cNa) for SAR calculation was determined from pNa measurement with Radelkis OP 113 type pX ion-selective electrode, according to $cNa = 10^{-pNa}$. Sodium Adsorption Ratio (SAR) is equal to $cNa / \sqrt{(cCa + cMg)}$.

$cMg)/2$), where ionic concentrations (c) are expressed in milliequivalents per liter measured in the water saturation extract of the soil. Assuming that there is negligible concentration of potassium, we can approximate $cCa + cMg$ as total cation concentration— cNa . As Richards [50] writes, electrical conductivity (expressed as $dS\ m^{-1}$) of saturation extract multiplied by 10 is approximately equal to the total cation concentration expressed in milliequivalents per liter. $EC\ [dS\ m^{-1}] * 10 - cNa \sim cCa + cMg$. With this approximation the SAR equation can be calculated. We should note that data on alkalinity, EC and SAR referring to the topsoil (0–30 cm) were used in further spatial modelling and assessment.

2.3. Environmental Covariates

The DEM derived from the UAV survey and a number of its derivatives (e.g., topographic position index, topographic wetness index, roughness) were used as environmental covariates in modelling the spatial distribution in the selected SAS indicators. The DEM derivatives listed in Table 1 were generated in SAGA GIS [51] with a spatial resolution of 2 m. Satellite-based spectral data and indices were generated as well using a custom script in Google Earth Engine to complete the UAV-derived DEM. Sentinel-2 images from the European Space Agency were filtered for the study site and time of field survey at first; additionally cloud masking was also applied to minimize cloud cover effect on the mean of six images meeting the mentioned filters. Covariate spectral indices (Table 1) were also calculated based on the Sentinel-2 bands: B2—blue, B3—green, B4—red, B8—near-infrared, B11—short-wavelength infrared 1, and B12—short-wavelength infrared 2. The final mean satellite image containing mentioned spectral bands and indices was exported in 2 m spatial resolution.

The reason for using the environmental covariates presented in Table 1 is partially explained in the introduction. Tóth et al. [11] showed that topography is one of the main influential factors determining the spatial variability in salt-affected soils and their indicators in the area of interest. Shrestha [52] used multiple regression analysis to examine the relationships between EC and spectral/soil properties and generated several models. They found that mid-infrared band (Landsat® band 7) and near-infrared (band 4) were found to be most correlated with the observed EC values in the surface soil layer. Nierd et al. [53], mapping natric soil areas, used the normalized difference ratio in Bands 5 and 4 with a threshold >0.19 . Most of the sites predicted to be natric were determined in the field to be natric (82%), but only half of the field-observed natric areas were correctly predicted. Dehni and Lounis [54] exploited the multi-spectral optical data from the LANDSAT ETM+ (Enhanced Thematic Mapper) to map surface states, including indices of salinity and sodicity as: BI: Brightness Index, NDSI: Normalized Difference Salinity Index, SI: Salinity Index, ASI: Aster Salinity Index (Agriculture), Index of Salinity (using GIS Geographic Information System and remote sensing), and finally the SSSI “Soil Salinity and Sodicity Index”. Furthermore, satellite images and derived spectral indices were used as they are proved to be highly informative covariates in predicting the small-scale variability in the SAS indicators [9].

Table 1. Summary of environmental covariates used for spatial modelling.

Factors of Soil Formation [10]	Environmental Covariates	Source
Soil surface	Brightness index	Sentinel-2
	Normalized difference salinity index	
	Salinity index 1-5	
	Salinity ratio	
	Visible infrared salinity index	
	Green band	
	Red band	
	Near-infrared band	
	Short-wave infrared-1	
	Short-wave infrared-2	

Table 1. Cont.

Factors of Soil Formation [10]	Environmental Covariates	Source
Topography	Elevation	DEM
	Slope	
	Aspect	
	Topographic position index	
	Terrain ruggedness index	
	Roughness	
	Flow direction	
	Catchment area	
	Modified catchment area	
	Diurnal anisotropic heating	
	LS factor (slope length factor)	
	Mass balance index	
	MRRTF	
Organisms	MRVBF	Sentinel-2
	Topographic wetness index	
	Normalized difference vegetation index	
	Soil adjusted vegetation index	
	Vegetation soil salinity index	

Legend: Abbreviations: DEM: digital elevation model, MRRTF: multiresolution index of ridge top flatness, and MRVBF: multiresolution index of valley bottom flatness.

2.4. Spatial Modelling and Predictive Mapping

A combination of machine learning algorithms and multivariate geostatistics, namely regression cokriging [9,55,56], was used for simultaneously predicting the spatial distribution in the selected SAS indicators. The reason for jointly modelling the spatial distribution of the indicators with multivariate geostatistics is that they can be spatially interdependent (or spatial cross-correlated) and, therefore, it is better to jointly model their spatial distribution. Taking the abovementioned into consideration, we adopted the following model to describe the spatial variations in the SAS indicators, which is frequently applied not just in digital soil mapping [57] but also in multivariate geostatistical modelling [58–60]:

$$\mathbf{Z}(\mathbf{u}) = \mathbf{m}(\mathbf{u}) + \boldsymbol{\varepsilon}(\mathbf{u})$$

$$\begin{bmatrix} Z_{\text{pH}}(\mathbf{u}) \\ Z_{\text{EC}}(\mathbf{u}) \\ Z_{\text{SAR}}(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} m_{\text{pH}}(\mathbf{u}) \\ m_{\text{EC}}(\mathbf{u}) \\ m_{\text{SAR}}(\mathbf{u}) \end{bmatrix} + \begin{bmatrix} \varepsilon_{\text{pH}}(\mathbf{u}) \\ \varepsilon_{\text{EC}}(\mathbf{u}) \\ \varepsilon_{\text{SAR}}(\mathbf{u}) \end{bmatrix} \quad (1)$$

where $\mathbf{Z}(\mathbf{u})$ is the vector of the SAS indicators (i.e., alkalinity, EC, and SAR), $\mathbf{m}(\mathbf{u})$ is the vector of the deterministic component describing the spatial variations in the SAS indicators that can be explained from the environmental covariates, $\boldsymbol{\varepsilon}(\mathbf{u})$ is the vector of the stochastic residuals that can be not just spatially correlated but spatially cross-correlated as well, and $\mathbf{u}$ is the vector of the geographical coordinates. Note that most kriging algorithms are optimal only if the variables are normally distributed [61]. Therefore, normal score transformation, which is a type of quantile transformation based on Gaussian anamorphosis [59,62], was performed on those SAS indicators whose distribution was found to be non-normal.

2.4.1. Ensemble Modelling for the Deterministic Component

The ensemble method we used consists of five independent learners, i.e., Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Neural Network (NN), and Generalized Linear Models with Lasso or Elastic Net Regularization (GLM), to model the deterministic component in Equation (1) for each SAS indicator. We used the R package ranger, which is a fast implementation of RF [63]. RF generates a large number of randomized and independent decision trees to predict and finally use all these predictions from each tree and aggregate them into one [64]. XGBoost is a method for tree boosting framework, which is highly applicable and widely employed in ML and works

based on the prediction error of the previous trees and creates a robust prediction [65]. SVM is another popular ML technique that assigns labels to each observation based on a new hyperplane and it can be used in both classification and regression problems [66,67]. NN utilizes multilayer feed-forward neural networks inspired by the human brain nervous system. This model includes sequentially connected multiple nodes in different layers and weights acquired by the network's knowledge [68]. Finally, the GLM technique is a generalized version of the linear regression model, the simplest and easiest predictive approach. The only assumption in this model is the existence of linear relations between observations and covariates, which is not the case in the real world. Here, this assumption is relaxed by computing the regularization path for the lasso or elastic-net penalty at a grid of values that can effectively manage the multicollinearity of covariates [69].

In this study, ensemble modeling was implemented by *train.splearner* function in the *Landmap* package [70] in R [71], which uses the SuperLearner method to stack all single learners. This package automates the spatial prediction mapping procedure by applying ensemble machine learning (using an extension of the *mlr* package [72]). This environment offers to run each model in parallel and combines all the individuals into one additional model (stacking). In this way, it takes the benefits of every single model, which increases the accuracy of mapping by combining the power of multiple models and reducing the variance and bias (bagging and boosting). Stacking, bagging, and boosting are the three main principles in applying ensembles [39]. In addition, *Landmap* package provides automation of several steps, such as deriving principal components, overlaying the observations and covariates, model fitting, hyper-parameter fine-tuning, and feature selection [73].

2.4.2. Multivariate Geostatistical Modelling of the Stochastic Residuals

Since our goal was to jointly model the spatial distribution of the SAS indicators (i.e., explicitly take their spatial interdependency into account), we used regression cokriging [9,55,56], where ensemble ML predictions for alkalinity, EC, and SAR were taken to be $m_{\text{pH}}(\mathbf{u})$, $m_{\text{EC}}(\mathbf{u})$, and $m_{\text{SAR}}(\mathbf{u})$ in Equation (1), respectively. For carrying out multivariate geostatistical modelling of $\epsilon(\mathbf{u})$ in Equation (1), we first derived the residuals, which means that we subtracted the ensemble ML predictions from the observed values for each SAS indicator at each sampling location. Next we computed the variograms and cross-variograms from the residuals and fitted a linear model of coregionalization (LMC) to ensure that a statistically valid model is applied [59].

We quantified the prediction uncertainty for each SAS indicator, where the uncertainty was identified by the kriging standard deviation with the assumption of normality and homoscedasticity [34]. For each SAS indicator, we compiled the 90% prediction interval map, which can be readily derived by subtracting and adding 1.64-times the kriging standard deviation to the prediction given by regression cokriging [74]. We should note that the quantified prediction uncertainty also allows for the compilation of probability maps (e.g., what is the probability that a given SAS indicator is not greater than a given threshold value), which is of great interest in practice.

As joint spatial modelling was performed on a normal scale, we had to transform the results back to the original scale for those SAS indicators where normal score transformation had been performed previously.

2.5. Validation

The performance of the spatial predictions for each SAS indicator was evaluated by 10-fold cross-validation in which the dataset was randomly partitioned into 10 equal-sized parts, then one of these parts was retained for validating the spatial predictions, which were obtained by using the remaining nine parts. This step was repeated until each of the 10 parts became a validation set exactly once. Four validation metrics were applied: (1) mean error (ME), which is also known as bias, (2) root mean square error (RMSE), which is the spread of the error distribution, (3) Lin's concordance correlation coefficient

(CCC; [75]), and (4) Nash–Sutcliffe model efficiency coefficient (NSE; [76]). These metrics can be computed as follows:

$$ME = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (3)$$

$$CCC = \frac{2 \sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sum_{i=1}^n (O_i - \bar{O})^2 + \sum_{i=1}^n (P_i - \bar{P})^2 + n(\bar{O} - \bar{P})^2} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (5)$$

where O_i and P_i are the observed and predicted SAS indicator value for the observation location i , respectively, and $\bar{O}$ and $\bar{P}$ are the mean of the predictions and observations.

The performance of uncertainty quantifications was evaluated by accuracy plots and G statistics. The theory behind an accuracy plot (also known as prediction interval coverage probability plot) is that if an uncertainty quantification reports, for example, a prediction interval with a 90% width, then we expect that 90% of the observations from the validation dataset fall within this prediction interval. This can be extended to symmetric prediction intervals with any width. Thus, an accuracy plot graphically presents what fraction of the observations from the validation dataset fall within symmetric prediction intervals with varying width. An accuracy plot ideally follows the $y = x$ line and the G statistics measures the overall closeness of the accuracy plot to this line:

$$G = 1 - \int_0^1 |\xi(p) - p| dp \quad (6)$$

where $\xi(p)$ and p are the fraction of the observations and the width of the prediction interval, respectively. Ideally, the value of G statistics is equal to 1.

3. Results

3.1. Soil Survey

Soils of the plot are not, or are slightly, saline (EC 0.14–0.43 dS m^{−1}; mean = 0.21 dS m^{−1}; Table 2) and alkaline (pH 7.9–8.8; mean = 8.2; Table 2) in the mapped topsoil. Sandy-silty texture is characteristic, the proportion of which depends on the location within the plot. Soil of formerly densely vegetated local depressions can be characterized by a higher silt fraction, organic matter, and salt content and, in addition, lower carbonate concentrations. Most of the measured soil properties show correlation with elevation (94.6–96.2 m AMSL). On the investigated plot, seven reference soil groups could be separated according to IUSS Working Group, World Reference Base for Soil Resources [49]: Chernozem (63.53%), Phaeozem (15.29%), Kastanozem (7.06%), Calcisol (5.88%), Gleysols (4.71%), Cambisols (2.35%), and Regosols (1.18%) [11]. There are no saline-sodic major soil groups among the soil profiles, as a result of agricultural use and management. Soil profiles have qualifiers (e.g., amphiprotosalic, endoprotosalic, katoprotosalic, protosodic), indicating that there are preserved signs of former salt and sodium accumulation. Salt maximums can be found in the C horizon in most cases according to ex situ measurements of the soil columns. Soil qualifiers support this observation. The horizon has high EC values (ECe ≥ 4 dS m^{−1}) e.g., appears mainly between 50 and 100 cm depth (endoprotosalic). Horizons, having more than 6% Na⁺ on the exchange complex, can be found, starting ≤100 cm from the soil surface (protosodic). Presence of the abovementioned soil groups represents the variation in SOC and thickness of the humus layer and alterations in the carbonate content and distribution and, furthermore, slight variation in water management (due to a channel near the plot and

periodically stagnant water in some parts of the study area). Mean of the calculated SAR values is 15.8 (Table 2), which refers to Na^+ dominance in soil. If the SAR value is above the threshold of 13, soil is considered to be sodic, having high levels of exchangeable sodium. It can cause soil structure deterioration due to soil particle dispersion, water infiltration problems, alkalinity, nutrient deficiencies, and plant toxicity (vines, beans, potatoes). The severity of problems due to soil sodicity depends on the texture, soil type, and drainage conditions. The abovementioned high SAR values appeared only in the northern part of the area, as shown in Figure 2.

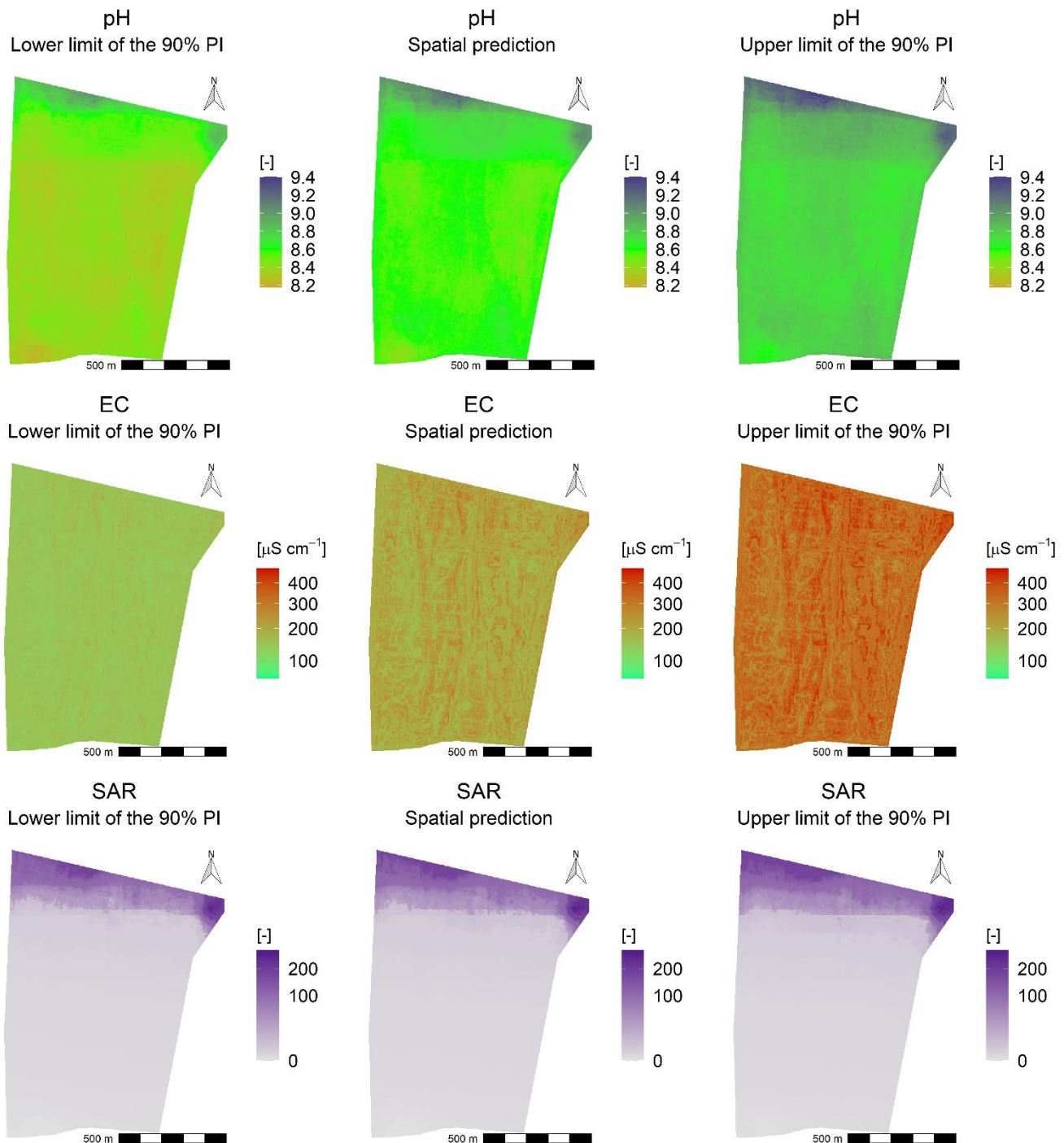


Figure 2. Spatial predictions (middle column) with the associated prediction uncertainty expressed by the 90% prediction interval (left and right column) for the indicators of salt-affected soils. Legend: Abbreviations: EC: electrical conductivity, SAR: sodium adsorption ratio, and PI: prediction interval.

Table 2. Summary statistics of the indicators of salt-affected soils computed from the point observations ($n = 85$).

SAS Indicators	Unit	Min	Max	Mean	SD
EC	$\mu\text{S cm}^{-1}$	136.4	428.0	214.0	59.94
pH	-	7.90	8.79	8.201	0.15
SAR	-	0.13	181.0	15.79	37.32

Legend: Abbreviations: SAS: salt-affected soils, EC: electrical conductivity, SAR: sodium adsorption ratio, and SD: standard deviation.

The groundwater table is shallow (160 to 301 cm depth with a mean of 200 cm). Groundwater is slightly alkaline–alkaline (pH 7.4–8.6 with a mean of 8.1) and saline (EC 2.4–6.1 dS m^{-1} with a mean of 4.18 dS m^{-1}), with Na^+ dominance (SAR 12.9–50.7 with a mean of 29.6), indicating saline-sodic alkaline conditions.

3.2. Spatial Modelling of SAS Indicators

The histogram of pH values showed normal distribution, which was also confirmed by Kolmogorov–Smirnov testing. The other two indicators (i.e., EC and SAR) did not follow the expected distribution (Figure 3). Thus, we had to apply normal score transformation for EC and SAR (Figure 3).

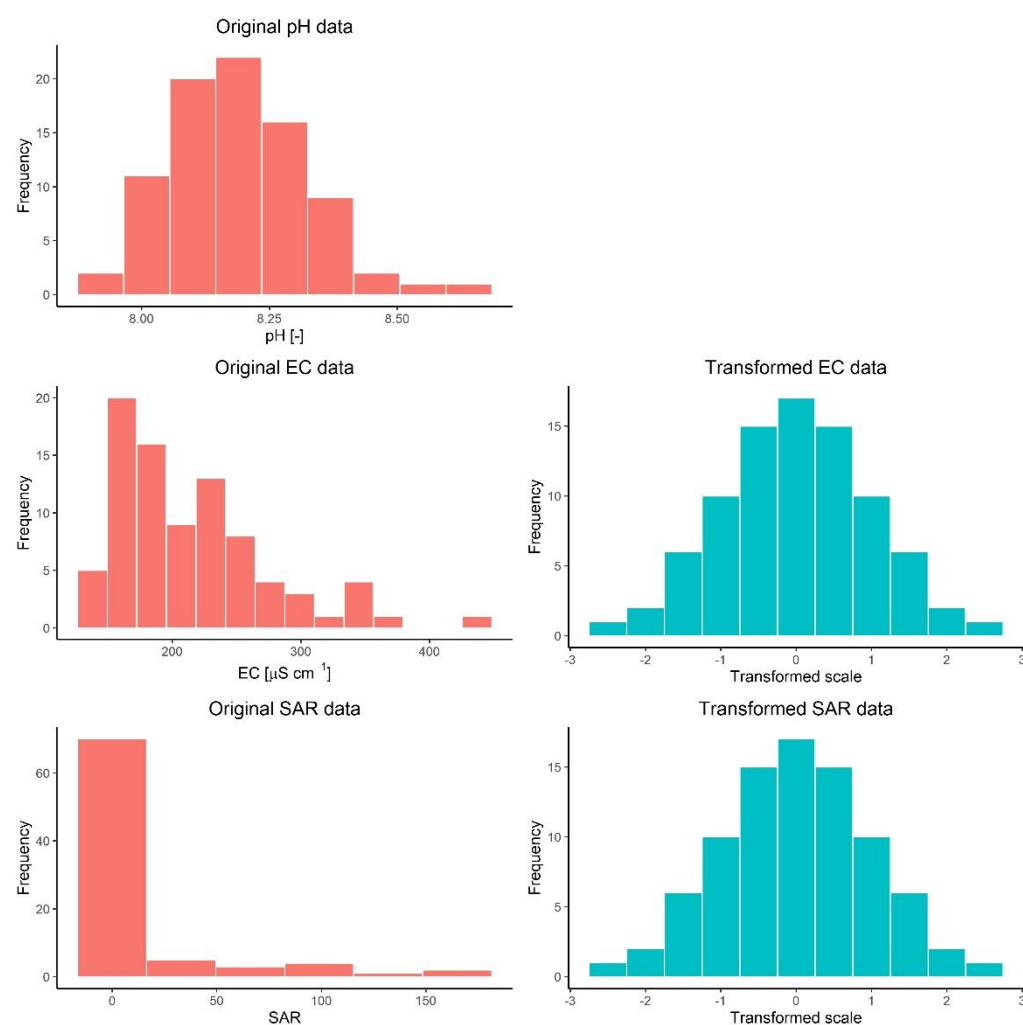


Figure 3. Histograms of the indicators of salt-affected soils before (left column) and after (right column) transformation. Legend: Abbreviations: EC: electrical conductivity and SAR: sodium adsorption ratio.

The results of machine learning are presented in Table 3. Having the highest accuracy (R^2) and least error (RMSE) in the SuperLearner model in predicting pH values and SAR is clear compared to the performance of each model individually. Regardless of the results of the SuperLearner, assessing the performance of each base learner indicated that RF was the best model related to the prediction of pH values and SAR, while the poorest performance for pH values was XGBoost and for SAR, was NN (Table 3). At the same time, spatial predictions of EC by SuperLearner and each of the sub-predictions by the mentioned models were not acceptable, except for RF. By plotting the spatial predictions of EC, the SuperLearner displayed artifacts and even spatial prediction of NN, and GLM showed irrational values, which is probably caused by some troubling covariates. To fix this problem in the case of EC, we only used RF, which showed favorable results and acceptable spatial prediction in the area. Therefore, in Table 3, we only reported the results for the RF model for EC. Additionally, when modelling the spatial variation in EC (Equation (1)) as one of the SAS indicators, RF prediction was taken to be $m_{EC}(\mathbf{u})$ in Equation (1). In the other cases, the SuperLearner prediction for pH values and SAR was taken to be $m_{pH}(\mathbf{u})$ and $m_{SAR}(\mathbf{u})$, respectively.

Table 3. Evaluation metrics summary for ensemble modelling.

ML Algorithms	R^2			RMSE			MAE		
	pH	EC	SAR	pH	EC	SAR	pH	EC	SAR
RF	0.36	0.39	0.96	0.12	0.97	0.21	0.10	0.79	0.09
XGBoost	0.09	-	0.91	5.42	-	0.83	5.42	-	0.65
NN	0.09	-	0.10	0.15	-	1.00	0.11	-	0.81
SVM	0.22	-	0.90	0.13	-	0.33	0.10	-	0.21
GLM	0.12	-	0.95	0.14	-	0.27	0.11	-	0.14
SuperLearner	0.43	-	0.96	0.11	-	0.20	0.09	-	0.11

Legend: Abbreviations: ML: machine learning, RMSE: root mean square error, MAE: mean absolute error, EC: electrical conductivity, SAR: sodium adsorption ratio, XGBoost: extreme gradient boosting, NN: neural network, SVM: support vector machine, and GLM: generalized linear model.

Direct and cross-variograms computed for the residuals of the ensemble models are presented in Figure 4 (open circles). For pH values and SAR, the residuals were derived by subtracting the SuperLearner predictions from the observed values of pH and SAR, while for EC, the residuals were computed by subtracting the RF predictions from the observed EC data. There is a clear spatial dependency and interdependency between the residuals, which confirms that it is reasonable to jointly model their spatial variability with multivariate geostatistics. To ensure that a statistically valid model is used in multivariate geostatistical modelling, we fitted a linear model of coregionalization [59], with a spherical model type and range value of 350 m (Figure 4, solid line).

The spatial prediction of SAS indicators with the prediction uncertainty (i.e., 90% prediction interval) is illustrated in Figure 2. As can be observed, soils with high values of pH and SAR were dominant in the northern parts of the study area, following the same pattern as our observations, proving these areas are affected by sodification/alkalinization. The map of alkalinity and SAR in other parts of the area showed minimal variations, with similar spatial patterns due to small changes in the topography of the area. Considering the uncertainty maps (Figure 2, right and left column), it is clear that the prediction uncertainty is higher in those parts where predictions have the highest values.

As mentioned in Section 2.4, an uncertainty model allows us to compile so-called probability maps, which are of great interest in practice. On the example of SAR, we compiled a map (Figure 5), which presents the probability that SAR is greater than the threshold value of 13. Taking this threshold value, the map actually shows the probability of finding sodic soils across the study site of interest. Based on the map, the northernmost part of the study site can be characterized by soils with relatively high sodium content, which is in line with our observations. This probability map can help stakeholders to calculate

the proportion of the plot area, which is entitled to receive statutory support. According to the threshold of the EU, based on a probability value of, e.g., 0.9, approximately 16.9% (0.16 km²) of the plot would meet the criterion for being subsidized.

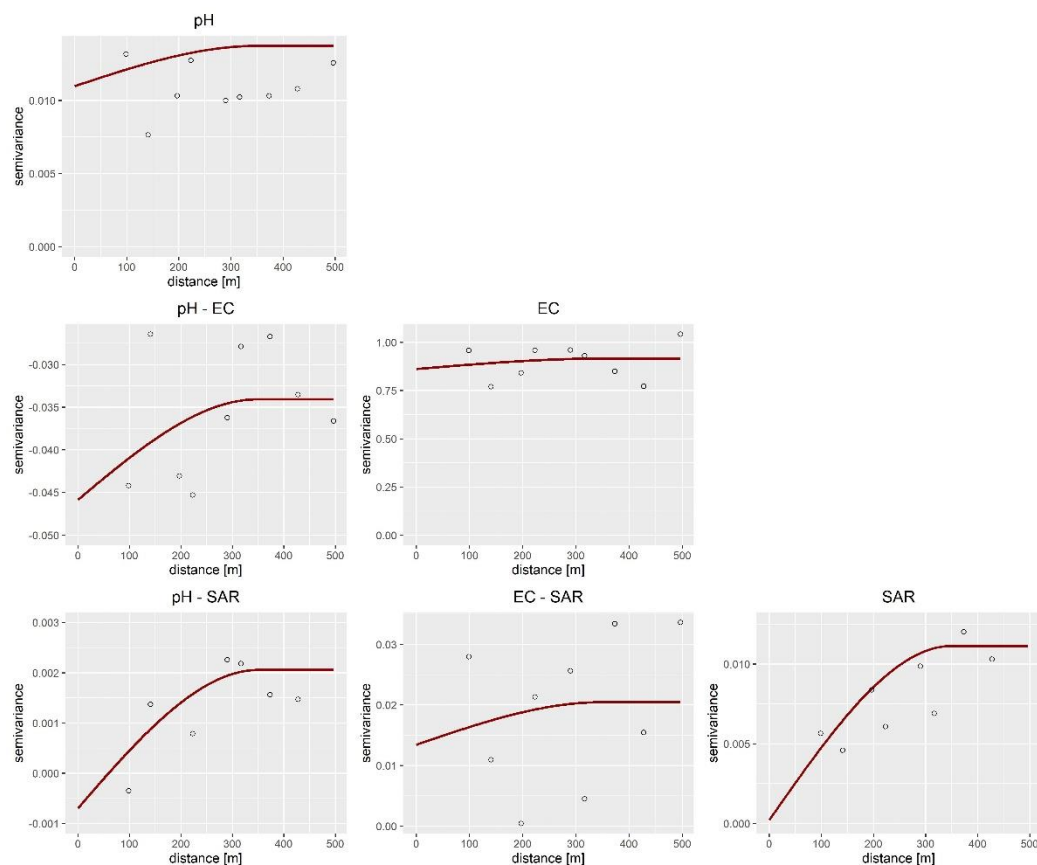


Figure 4. The computed direct and cross-variograms (open circles) and fitted linear model of coregonalization (solid line).

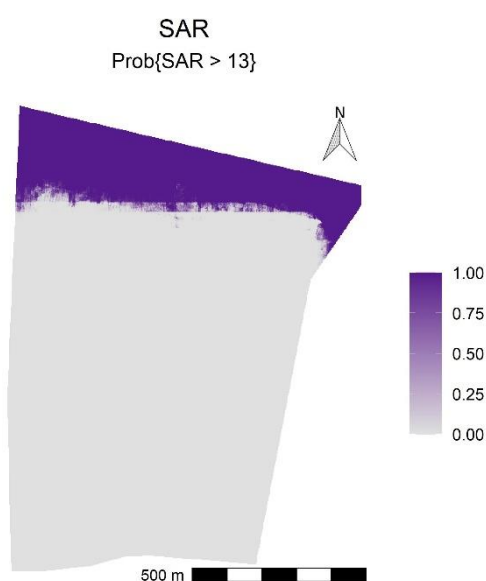


Figure 5. Map on sodium adsorption ratio (SAR) showing the probability that SAR is greater than the threshold value of 13.

3.3. Performance of Spatial Predictions and Uncertainty Quantifications

As mentioned, we performed 10-fold cross-validation to validate the results of spatial prediction presented in Figure 2 (middle column). Therefore, the computed values of ME, RMSE, CCC, and NSE are reported in Table 4. The best unbiased results can be diagnosed from the smallest possible ME and RMSE while having the highest accuracy. Based on our results, our spatial predictions were acceptable in this case. CCC values varied from 0.39 to 0.97. In addition, NSE, which is one of the model efficiency measurements, can be equivalent to the value of R square in the application of regression procedures if its value is greater than zero. Here, the NSE values varied from the highest predictive skill for SAR (NSE = 0.97) to the lowest for EC (NSE = 0.24).

Table 4. The performance of spatial predictions by 10-fold cross-validation.

SAS Indicators	ME	RMSE	CCC	NSE
pH	0.001	0.11	0.59	0.41
EC	0.001	0.86	0.39	0.24
SAR	0.007	0.22	0.97	0.95

Legend: Abbreviations: SAS: salt-affected soils, EC: electrical conductivity, SAR: sodium adsorption ratio, ME: mean error, RMSE: root mean square error, CCC: Lin's concordance correlation coefficient, and NSE: Nash-Sutcliffe model efficiency coefficient.

In Figure 6, the compiled accuracy plot with the computed G statistic for each SAS indicator is presented. The accuracy plots follow the $y = x$ line, meaning that the uncertainty quantifications are accurate for each indicator. This is also confirmed by the computed G statistics, which is close to its expected value (i.e., 1) in each case.

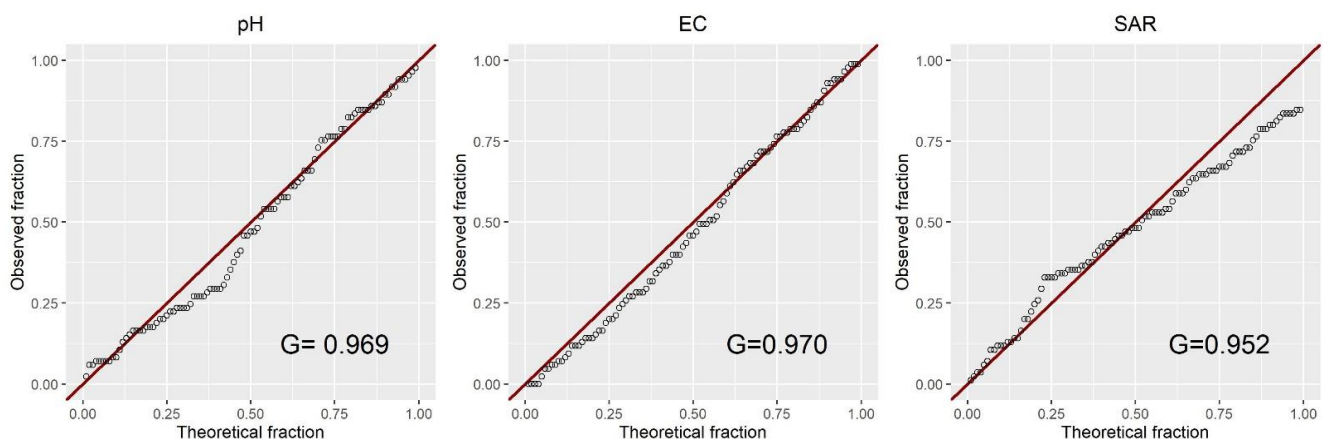


Figure 6. Accuracy plots with the computed G statistics. Legend: Abbreviations: EC: electrical conductivity, and SAR: sodium adsorption ratio.

4. Discussion

4.1. Ensemble Machine Learning

We found that the SuperLearner outperformed each base learner (i.e., RF, XGBoost, SVM, NN, and GLM) in predicting pH values and SAR (Table 3), which can be explained by the fact that ensemble modelling integrates the individual models and can reduce noise and variance in prediction while avoiding overfitting, which, altogether, leads to higher and more robust performance than any other single model. This finding is in line with the international literature. Several studies have demonstrated the effectiveness of ensemble modelling over single modelling [60–62]. Additionally, Mishra et al. [30] applied different machine learning approaches with regression kriging on the example of soil organic carbon. They found that an ensemble prediction approach is better than any individual model while providing greater spatial details and higher accuracy for predicting the spatial variation in the soil property of interest. However, we should note that RF

was the best-performing model among the five base learners, showing highly comparable results and almost identical to the SuperLearner. This could be attributed to the capability of RF to handle the non-linearity of data and outliers [30].

Although the result of this study confirmed a better performance of the SuperLearner for pH values and SAR, in the case of EC, RF outperformed the SuperLearner and, therefore, RF was used to model the deterministic component for EC (i.e., $m_{EC}(\mathbf{u})$ in Equation (1)) instead of the SuperLearner. The outstanding performance of RF, in this case, can be attributed to several reasons, e.g., the nature of EC itself, which is usually a soil property that changes quickly over space and time at the field scale [77,78], the presence of artifacts in covariates, insufficient covariates to explain the EC spatial variation in the study area, and the number of observations that might be few for ensemble modelling [79].

In general, when applying any spatial prediction model, we have to consider that the ability to predict will depend on multiple reasons, e.g., the relation between the soil attribute and environmental covariates [80,81], sampling accuracy of field observations, type of environmental covariates, and their resolution [60]. Accordingly, each model for each property has a different capability to predict in every region. Therefore, no most-favorable model could be applied to all situations. Multiple single models and ensembles of these models should be assessed since the ensemble will use the capability of each model and, in general, can boost their prediction and accuracy.

4.2. Multivariate Geostatistics

The computed direct and cross-variograms presented in Figure 4 confirmed our hypothesis, i.e., the SAS indicators across the area of interest are spatially interdependent (i.e., spatially cross-correlated) and, therefore, it is better to jointly model their spatial distribution using multivariate geostatistics. The application of multivariate geostatistics in digital soil mapping is well established [82–84] and, recently, Szatmári et al. [9] produced a detailed discussion on its pros and cons in SAS mapping. In this study, its most important added value is that not just the spatial predictions of the SAS indicators but also their prediction uncertainty is connected by the spatial cross-correlation existing between the indicators, which has a lot of merits, especially if a complex assessment of the indicators is targeted, for instance, in precision agriculture or farm-scale planning.

We should note that some of the computed variograms showed relatively large nugget variance (i.e., discontinuity at the origin; see Figure 4), which is not rare in digital soil mapping (e.g., [85]). Concerning our study, the large nugget variance can be mainly attributed to the sampling strategy we applied. Although systematic grid sampling is quasi optimal for soil-mapping purposes [86–89], it is not suitable for exploring small-scale variability [90–92]; to be more precise, spatial variability here is shorter than the distance between two neighboring sampling locations.

4.3. Mapping and Assessment of Salt-Affected Soils

One of the most characteristic features of the Hungarian lowlands is salt accumulation and the related processes of sodification and alkalinization [93]. Nowadays, as have to produce more and more food due to the growing population, even less suitable areas, such as SAS, are brought under agricultural production. From an agricultural cultivation point of view, it is essential to create maps in an appropriately high resolution in meliorated saline areas, in order not only to support precision agricultural production for better yields, but to highly represent SAS indicators—as limiting factors for cultivation. For this, it is necessary to create additional maps that inform about the probability of the appearance of the given limiting factor (SAR value in this case). This is important not only for the planning of the agrotechnical operation to be selected, but it also informs about the amount of support that can be requested for the given agricultural land and the size of the area to be supported. There are several mechanisms that provide incentives for farmers that crop salt-affected soils; one is provided by the EU [94]. The EU links subsidies to threshold values of salinity ($EC_e > 4 \text{ dS} \cdot \text{m}^{-1}$) and sodicity ($ESP > 6$; $SAR > 13$). According to these

thresholds, based on a probability value of 0.9, approximately 16.9% (0.16 km²) of the plot would meet the criterion for being subsidized. Soil sodicity was not fully delineated because of a lack of suitable data at lower salt levels, which are still limiting for cultivation. When salinity is large, its mapping is rather straightforward and our approach shows the case of slight salinization, a borderline case when mapping is more complicated.

Salinization is important not only on hayfields and grazelands, but it is the most abundant form of chemical degradation for arable lands [95]. For now, salinity values decreased to a medium or low category in the topsoil (ploughed horizon) of the sampling plot, as a result of agrotechnical operations. Salt maximums can be found mainly in the C horizon. The spatial distribution of topsoil EC is varied, following the topography (Figure 2), as salinity shows correlation with elevation. Nabiollahi et al. [96] also assessed SAS indicators using digital mapping and hybridized random forests and found that the most important covariates to predict pH, EC, and SAR were groundwater table, categorical maps, salinity index, and multi-resolution ridge top flatness (MRRTF). This supports that, in addition to topography and elevation, groundwater table is crucial for appropriate assessment of SAS indicators. Groundwater table from ten locations in our sampling area ranged from 160 to 301 cm depth, with 2.4–6.1 dS·m^{−1} salt content, which is about the critical groundwater level (according to Kovda [97]), carrying the risk of salt accumulation in the fluctuation zone at the groundwater level. For the sake of the small number of sample points, information on groundwater table was not available spatially exhaustively in our sampling site; therefore, we could not take it into consideration in the course of spatial prediction. However, as irrigation is not applied and the number and length of dry periods are increasing [98], negative water balance (enhanced soil evaporation and upward movement of soil moisture) might also cause further and repeated salinity problems in the ploughed horizon in the future. Soil salinity is a fast-changing parameter; furthermore, depending on the weather conditions, there is a risk of secondary salt accumulation. Detailed and recurring mapping of soil salinity is necessary on previously saline, meliorated agricultural areas.

More alkaline soils (pH > 8.5) can be found in the northern part of the study site (Figure 2.), as well as higher SAR values. Sparks [99] notes that, in soils with high electrolyte concentrations, pH value is usually < 8.5 and the soil is flocculated. However, if the soluble salts are leached out, usually Na⁺ becomes an even greater problem and the soil pH rises to > 8.5 and the soil can become dispersed [50]. This phenomenon can cause waterlogging in the topsoil and impede plowing and agrotechnical operations involving soil loosening. Furthermore, alkalinity inhibits a number of essential plant nutrients' uptake for crops, influencing both the availability of soil nutrients and affecting losses of certain nutrients from the soil system, reducing the expected yield. After delimiting these alkaline parts of the area on a high-resolution digital map, the stakeholder will be able to plan cost-effective interventions.

5. Conclusions

The aim of this study was to map and assess the salt-affectedness on an area of arable land in Hungary, with high spatial resolution, using a combination of ensemble machine learning and multivariate geostatistics on three SAS indicators (i.e., alkalinity, EC, and SAR). In ensemble modelling, five base learners were used, namely Ranger, XGBoost, SVM, NN, and GLM. Our results illustrated that ensemble machine learning combined with multivariate geostatistics could be a promising method not just for jointly modelling and mapping the spatial distribution of different indicators of salt-affected soils at high spatial resolution but also in assessing salt-affectedness on arable lands at the field scale. However, the results also showed that ensemble machine learning does not necessarily perform better than the base learners separately. In such a case, it is better not to use ensemble modelling and instead use the best base learner alone. In Hungary, there were several mapping campaigns of salt-affected soils; however, the salinity/sodicity/alkalinity of soils are continuously present in the landscape, showing specific depth distribution and intensity

in accordance with geomorphology and hydraulic conditions. These SAS indicators need to be mapped more precisely to help in the accurate execution of new precision agricultural operations. High-spatial-resolution mapping is important for stakeholders:

1. For farmers to indicate:
 - The part of the plot where reclamation/drainage works must be carried out and
 - Site-specific cultivation can be managed;
2. For decision-makers to help them decide:
 - the allocation of subsidies.

Author Contributions: Conceptualization, T.T. and L.P.; methodology, F.H., K.B., J.M., M.Á. and G.S.; software, F.H., J.M., M.Á. and G.S.; validation, F.H., K.B., T.T., J.M., M.Á., T.J.N., L.P. and G.S.; formal analysis, F.H., T.T., J.M., L.P. and G.S.; investigation, J.M., M.Á., S.K. and P.L.; resources, J.M., M.Á., S.K. and P.L.; data curation, J.M., M.Á., Z.A.K. and N.S.-V.; writing—original draft preparation, F.H., K.B., J.M., M.Á. and G.S.; writing—review and editing, K.B., L.P. and G.S.; visualization, F.H., M.Á. and G.S.; supervision, T.T. and L.P.; project administration, K.B.; funding acquisition, T.T. and L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research, Development and Innovation Office (NKFIH), grant numbers K-131820 and K-124290. The work of G.S. was supported by the Premium Postdoctoral Scholarship of the Hungarian Academy of Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are deeply indebted to the following persons: Péter Horváth (Dunavecsei Mg. Zrt.) and István Fekete for facilitating the field work in the investigated plot; Márton Tóth for helping in the description of profiles and for assisting in the field and laboratory work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Stolte, J.; Tesfai, M.; Oygarden, L.; Kvaerno, S.; Keizer, J.; Verheijen, F.; Panagos, P.; Ballabio, C.; Hessel, R. Soil Threats in Europe: Status, Methods, Drivers and Effects on Ecosystem Services. 2016. Available online: <https://esdac.jrc.ec.europa.eu/content/soil-threats-europe-status-methods-drivers-and-effects-ecosystem-services> (accessed on 1 June 2022).
2. Daliakopoulos, I.N.; Tsanis, I.K.; Koutroulis, A.; Kourgialas, N.N.; Varouchakis, A.E.; Karatzas, G.P.; Ritsema, C.J. The Threat of Soil Salinity: A European Scale Review. *Sci. Total Environ.* **2016**, *573*, 727–739. [CrossRef] [PubMed]
3. Majeed, A.; Muhammad, Z. Salinity: A Major Agricultural Problem—Causes, Impacts on Crop Productivity and Management Strategies. *Plant Abiotic Stress Toler. Agron. Mol. Biotechnol. Approaches* **2019**, 83–99. [CrossRef]
4. Huang, J.; Hartemink, A.E. Soil and Environmental Issues in Sandy Soils. *Earth-Sci. Rev.* **2020**, *208*, 103295. [CrossRef]
5. Singh, A. Soil Salinity: A Global Threat to Sustainable Development. *Soil Use Manag.* **2022**, *38*, 39–67. [CrossRef]
6. Tomaz, A.; Palma, P.; Alvarenga, P.; Gonçalves, M.C. Soil Salinity Risk in a Climate Change Scenario and Its Effect on Crop Yield. *Clim. Chang. Soil Interact.* **2020**, 351–396. [CrossRef]
7. Ferreira, C.S.S.; Seifollahi-Aghmiuni, S.; Destouni, G.; Ghajarnia, N.; Kalantari, Z. Soil Degradation in the European Mediterranean Region: Processes, Status and Consequences. *Sci. Total Environ.* **2022**, *805*, 150106. [CrossRef]
8. Omuto, C.T.; Vargas, R.R.; El Mobarak, A.M.; Mohamed, N.; Viatkin, K.; Yigini, Y. *Mapping of Salt-Affected Soils: Technical Manual*; FAO: Rome, Italy, 2020.
9. Szatmári, G.; Bakacsi, Z.; Laborczi, A.; Petrik, O.; Pataki, R.; Tóth, T.; Pásztor, L. Elaborating Hungarian Segment of the Global Map of Salt-Affected Soils (GSSmap): National Contribution to an International Initiative. *Remote Sens.* **2020**, *12*, 4073. [CrossRef]
10. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; McGraw-Hill: New York, NY, USA, 1941.
11. Tóth, T.; Gallai, B.; Novák, T.; Czigány, S.; Makó, A.; Kocsis, M.; Árvai, M.; Mészáros, J.; László, P.; Koós, S.; et al. Practical Evaluation of Four Classification Levels of Soil Taxonomy, Hungarian Classification and WRB in Terms of Biomass Production in a Salt-Affected Alluvial Plot. *Geoderma* **2022**, *410*, 115666. [CrossRef]
12. Terres, J.; Toth, T.; Wania, A.; Hagyo, A.; Koeble, R.; Nisini Scacchiachichi, L. *Updated Guidelines for Applying Common Criteria to Identify Agricultural Areas with Natural Constraints*; JRC: Tokyo, Japan, 2016.
13. Szabolcs, I. Review of Research on Salt Affected Soils. *Nat. Resour. Res.* **1979**. [CrossRef]
14. Sigmond, E. Hungarian Alkali Soils and Methods of Their Reclamation. In *Special Publication Issued by the California Agricultural Experiment Station*; University of California Printing Office: Berkeley, CA, USA, 1927.

15. Szabolcs, I. *Salt-Affected Soils*; CRC Press: Boca Raton, FL, USA, 1989.
16. Szabolcs, I. *European Solonchaks Soils and Their Reclamation*; Akadémia Kiadó: Budapest, Hungary, 1971.
17. Tóth, T.; Szendrei, G. A Hazai Szikes Talajok És a Szikesedés Valamint a Sófelhalmozódási Folyamatok Rövid Jellemzése. *Topogr. Mineral. Hungariae* **2006**, *9*, 7–20.
18. Mádlné Szőnyi, J.; Simon, S.; Tóth, J.; Pogácsás, G. Connection between Surface and Groundwaters in the Case of Kelemen-Lake and Kolon-Lake. *Általános Földtani Szle.* **2005**, *30*, 93–110.
19. Bakacsi, Z.; Kuti, L. Agrogeological Investigation on a Salt Affected Landscape in the Danube Valley, Hungary. *Agrokémia és Talajt.* **1998**, *47*, 29–38.
20. Arany, S. A Hortobágyi Ősi Szikes Legelőkön Végzett Talajfelvételek. *Kísérletügyi Közlemények Pallas részvénytársaság sajtója* **1926**, *29*, 48–71.
21. Magyar, P. Adatok a Hortobágy Növényészeti és Geobotanikai Viszonyaihoz. *Erdészeti kísérletek* **1928**, *30*, 26–63.
22. Szabolcs, I. *Hortobágy Talajai*; Mezőgazdasági Kiadó: Budapest, Hungary, 1954.
23. Minasny, B.; McBratney, A.B. Digital Soil Mapping: A Brief History and Some Lessons. *Geoderma* **2016**, *264*, 301–311. [\[CrossRef\]](#)
24. Lagacherie, P.; McBratney, A.B. Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. *Dev. Soil Sci.* **2007**, *31*, 3–22.
25. Szatmári, G.; Pásztor, L.; Heuvelink, G.B.M. Estimating Soil Organic Carbon Stock Change at Multiple Scales Using Machine Learning and Multivariate Geostatistics. *Geoderma* **2021**, *403*, 115356. [\[CrossRef\]](#)
26. Heuvelink, G.B.M.; Angelini, M.E.; Poggio, L.; Bai, Z.; Batjes, N.H.; van den Bosch, R.; Bossio, D.; Estella, S.; Lehmann, J.; Olmedo, G.F.; et al. Machine Learning in Space and Time for Modelling Soil Organic Carbon Change. *Eur. J. Soil Sci.* **2020**. [\[CrossRef\]](#)
27. Styc, Q.; Lagacherie, P. Uncertainty Assessment of Soil Available Water Capacity Using Error Propagation: A Test in Languedoc-Roussillon. *Geoderma* **2021**, *391*, 114968. [\[CrossRef\]](#)
28. Helfenstein, A.; Mulder, V.L.; Heuvelink, G.B.M.; Okx, J.P. Tier 4 Maps of Soil PH at 25 m Resolution for the Netherlands. *Geoderma* **2022**, *410*, 115659. [\[CrossRef\]](#)
29. Chen, S.; Mulder, V.L.; Martin, M.P.; Walter, C.; Lacoste, M.; Richer-de-Forges, A.C.; Saby, N.P.A.; Loiseau, T.; Hu, B.; Arrouays, D. Probability Mapping of Soil Thickness by Random Survival Forest at a National Scale. *Geoderma* **2019**, *344*, 184–194. [\[CrossRef\]](#)
30. Keskin, H.; Grunwald, S. Regression Kriging as a Workhorse in the Digital Soil Mapper's Toolbox. *Geoderma* **2018**, *326*, 22–41. [\[CrossRef\]](#)
31. Sahbeni, G.; Székely, B. Spatial Modeling of Soil Salinity Using Kriging Interpolation Techniques: A Study Case in the Great Hungarian Plain. *Eurasian J. Soil Sci.* **2022**, *11*, 102–112. [\[CrossRef\]](#)
32. Brus, D.J.; Kempen, B.; Heuvelink, G.B.M. Sampling for Validation of Digital Soil Maps. *Eur. J. Soil Sci.* **2011**, *62*, 394–407. [\[CrossRef\]](#)
33. Heuvelink, G.B.M. Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling. In *Pedometrics*; Springer: Cham, Germany, 2018; pp. 439–461.
34. Szatmári, G.; Pásztor, L. Comparison of Various Uncertainty Modelling Approaches Based on Geostatistics and Machine Learning Algorithms. *Geoderma* **2019**, *337*, 1329–1340. [\[CrossRef\]](#)
35. Mishra, U.; Gautam, S.; Riley, W.J.; Hoffman, F.M. Ensemble Machine Learning Approach Improves Predicted Spatial Variation of Surface Soil Organic Carbon Stocks in Data-Limited Northern Circumpolar Region. *Front. Big Data* **2020**, *3*, 528441. [\[CrossRef\]](#)
36. Brungard, C.; Nauman, T.; Duniway, M.; Veblen, K.; Nehring, K.; White, D.; Salley, S.; Anchang, J. Regional Ensemble Modeling Reduces Uncertainty for Digital Soil Mapping. *Geoderma* **2021**, *397*, 114998. [\[CrossRef\]](#)
37. Kadavi, P.R.; Lee, C.W.; Lee, S. Application of Ensemble-Based Machine Learning Models to Landslide Susceptibility Mapping. *Remote Sens.* **2018**, *10*, 1252. [\[CrossRef\]](#)
38. Hengl, T.; Miller, M.A.E.; Križan, J.; Shepherd, K.D.; Sila, A.; Kilibarda, M.; Antonijević, O.; Glušica, L.; Dobermann, A.; Haefele, S.M.; et al. African Soil Properties and Nutrients Mapped at 30 m Spatial Resolution Using Two-Scale Ensemble Machine Learning. *Sci. Reports* **2021**, *11*, 1–18. [\[CrossRef\]](#)
39. Zhang, C.; Ma, Y. *Ensemble Machine Learning*; Springer: New York, NY, USA, 2012.
40. Seni, G.; Elder, J.F. *Ensemble Methods in Data Mining*. Springer International Publishing, 2010; Available online: <https://link.springer.com/content/pdf/bfm%3A978-3-031-01899-2%2F1> (accessed on 1 June 2022).
41. Polikar, R. Ensemble Learning. *Ensemble Mach. Learn.* **2012**, 1–34. [\[CrossRef\]](#)
42. El Badawi, H.; Azais, F.; Bernard, S.; Comte, M.; Kerzerho, V.; Lefevre, F. Use of Ensemble Methods for Indirect Test of RF Circuits: Can It Bring Benefits? In Proceedings of the 2019 IEEE Latin American Test Symposium (LATS), Santiago, Chile, 11–13 March 2019. [\[CrossRef\]](#)
43. Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The Use of Remote Sensing in Soil and Terrain Mapping—A Review. *Geoderma* **2011**, *162*, 1–19. [\[CrossRef\]](#)
44. Dwivedi, R.S. *Remote Sensing of Soils*; Springer: Berlin/Heidelberg, Germany, 2017; ISBN 9783662537404.
45. Pásztor, L.; Takács, K. Remote Sensing in Soil Mapping—A Review. *Agrokémia és Talajt.* **2014**, *63*, 353–370. [\[CrossRef\]](#)
46. Wang, N.; Peng, J.; Chen, S.; Huang, J.; Li, H.; Biswas, A.; He, Y.; Shi, Z. Improving Remote Sensing of Salinity on Topsoil with Crop Residues Using Novel Indices of Optical and Microwave Bands. *Geoderma* **2022**, *422*, 115935. [\[CrossRef\]](#)
47. Sahbeni, G. A PLSR Model to Predict Soil Salinity Using Sentinel-2 MSI Data. *Open Geosci.* **2021**, *13*, 977–987. [\[CrossRef\]](#)

48. Dövényi, Z. *Inventory of Microregions in Hungary*, 2nd ed.; Hungarian Academy of Sciences, Geographical Research Institute: Budapest, Hungary, 2010.
49. IUSS Working Group. *WRB World Reference Base for Soil Resources 2014, Update 2015 International Soil Classification System for Naming Soils and Creating Legends for Soil Maps*; World Soil Resources Reports No. 106; IUSS Working Group: Rome, Italy, 2015.
50. Richards, L.A. *Diagnosis and Improvement of Saline Alkali Soils*; US Department of Agriculture: Washington, DC, USA, 1954.
51. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [\[CrossRef\]](#)
52. Shrestha, R.P. Relating Soil Electrical Conductivity to Remote Sensing and Other Soil Properties for Assessing Soil Salinity in Northeast Thailand. *Land Degrad. Dev.* **2006**, *17*, 677–689. [\[CrossRef\]](#)
53. Nield, S.J.; Boettinger, J.L.; Ramsey, R.D. Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data. *Soil Sci. Soc. Am. J.* **2007**, *71*, 245–252. [\[CrossRef\]](#)
54. Dehni, A.; Lounis, M. Remote Sensing Techniques for Salt Affected Soil Mapping: Application to the Oran Region of Algeria. *Procedia Eng.* **2012**, *33*, 188–198. [\[CrossRef\]](#)
55. Stein, A.; Corsten, L.C.A. Universal Kriging and Cokriging as a Regression Procedure. *Biometrics* **1991**, *47*, 575–587. [\[CrossRef\]](#)
56. Myers, D.E. Matrix Formulation of Co-Kriging. *J. Int. Assoc. Math. Geol.* **1982**, *14*, 249–257. [\[CrossRef\]](#)
57. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52. [\[CrossRef\]](#)
58. Wackernagel, H. *Multivariate Geostatistics*; Springer: Berlin/Heidelberg, Germany, 2003.
59. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA; Oxford, UK, 1997; ISBN 9780195115383.
60. Cressie, N.A.C. *Statistics for Spatial Data*; Wiley: Hoboken, NJ, USA, 1993.
61. Geiger, J. Some Thoughts on the Pre- and Post-Processing in Sequential Gaussian Simulation and Their Effects on Reservoir Characterization. In *New Horizons in Central European Geomathematics, Geostatistics and Geoinformatics*; Geiger, J., Pál-Molnár, E., Malvic, T., Eds.; GeoLittera: Szeged, Hungary, 2012; pp. 17–34.
62. Deutsch, C.V.; Journel, A.G. *GSLIB: Geostatistical Software Library and User's Guide*; Oxford University Press: New York, UK, 1998; ISBN 9780195100150.
63. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
64. Wright, M.N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [\[CrossRef\]](#)
65. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [\[CrossRef\]](#)
66. Pisner, D.A.; Schnyer, D.M. Support Vector Machine. *Mach. Learn. Methods Appl. Brain Disord.* **2020**, 101–121. [\[CrossRef\]](#)
67. Vapnik, V.; Golowich, S.; Smolal, A. Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. *Adv. Neural Inf. Processing Syst.* **1996**, *9*. [\[CrossRef\]](#)
68. Zhu, A.X. Mapping Soil Landscape as Spatial Continua: The Neural Network Approach. *Water Resour. Res.* **2000**, *36*, 663–677. [\[CrossRef\]](#)
69. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1. [\[CrossRef\]](#)
70. Hengl, T. Landmap. 2021.
71. R Core Team. *A Language and Environment for Statistical Computing*; The R Foundation—The R Project for Statistical Computing: Vienna, Austria, 2013. Available online: <http://www.R-project.org/> (accessed on 1 June 2022).
72. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. Mlr: Machine Learning in R. *J. Mach. Learn. Res.* **2016**, *17*, 5938–5942.
73. Hengl, T.; Parente, L.; Bonannella, C. Spatial and Spatiotemporal Interpolation/Prediction Using Ensemble Machine Learning. 2022. Available online: <https://opengeohub.github.io/spatial-prediction-emi/> (accessed on 1 June 2022). [\[CrossRef\]](#)
74. Heuvelink, G. Uncertainty Quantification of GlobalSoilMap Products. In *GlobalSoilMap*; CRC Press: Boca Raton, FL, USA, 2014; pp. 335–340.
75. Lin, L.I.-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **1989**, *45*, 255–268. [\[CrossRef\]](#)
76. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [\[CrossRef\]](#)
77. Paz, A.M.; Castanheira, N.; Farzamian, M.; Paz, M.C.; Gonçalves, M.C.; Monteiro Santos, F.A.; Triantafyllis, J. Prediction of Soil Salinity and Sodicity Using Electromagnetic Conductivity Imaging. *Geoderma* **2020**, *361*, 114086. [\[CrossRef\]](#)
78. Li, H.Y.; Shi, Z.; Webster, R.; Triantafyllis, J. Mapping the Three-Dimensional Variation of Soil Salinity in a Rice-Paddy Soil. *Geoderma* **2013**, *195*, 31–41. [\[CrossRef\]](#)
79. Lück, E.; Gebbers, R.; Ruehlmann, J.; Spangenberg, U. Electrical Conductivity Mapping for Precision Farming. *Near Surf. Geophys.* **2009**, *7*, 15–26. [\[CrossRef\]](#)
80. Hateffard, F.; Dolati, P.; Heidari, A.; Zolfaghari, A.A. Assessing the Performance of Decision Tree and Neural Network Models in Mapping Soil Properties. *J. Mt. Sci.* **2019**, *16*, 1833–1847. [\[CrossRef\]](#)

81. John, K.; Isong, I.A.; Kebonye, N.M.; Ayito, E.O.; Agyeman, P.C.; Afu, S.M. Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land* **2020**, *9*, 487. [CrossRef]
82. Odeh, I.O.A.; McBratney, A.B.; Chittleborough, D.J. Further Results on Prediction of Soil Properties from Terrain Attributes: Heterotopic Cokriging and Regression-Kriging. *Geoderma* **1995**, *67*, 215–226. [CrossRef]
83. Lark, R.M.; Ander, E.L.; Cave, M.R.; Knights, K.V.; Glennon, M.M.; Scanlon, R.P. Mapping Trace Element Deficiency by Cokriging from Regional Geochemical Soil Data: A Case Study on Cobalt for Grazing Sheep in Ireland. *Geoderma* **2014**, *226–227*, 64–78. [CrossRef]
84. Wackernagel, H. Cokriging versus Kriging in Regionalized Multivariate Data Analysis. *Geoderma* **1994**, *62*, 83–92. [CrossRef]
85. Vaysse, K.; Lagacherie, P. Evaluating Digital Soil Mapping Approaches for Mapping GlobalSoilMap Soil Properties from Legacy Data in Languedoc-Roussillon (France). *Geoderma Reg.* **2015**, *4*, 20–30. [CrossRef]
86. Szatmári, G.; Barta, K.; Pásztor, L. An Application of a Spatial Simulated Annealing Sampling Optimization Algorithm to Support Digital Soil Mapping. *Hungarian Geogr. Bull.* **2015**, *64*, 35–48. [CrossRef]
87. Brus, D.J. Sampling for Digital Soil Mapping: A Tutorial Supported by R Scripts. *Geoderma* **2019**, *338*, 464–480. [CrossRef]
88. Burgess, T.M.; Webster, R.; McBratney, A.B. Optimal Interpolation and Isarithmic Mapping of Soil Properties. IV Sampling Strategy. *J. Soil Sci.* **1981**, *32*, 643–659. [CrossRef]
89. de Gruijter, J.J.; Bierkens, M.F.P.; Brus, D.J.; Knotters, M. *Sampling for Natural Resource Monitoring*; Springer: Berlin/Heidelberg, Germany, 2006; ISBN 978-3-540-22486-0.
90. Marchant, B.P.; Lark, R.M. Optimized Sample Schemes for Geostatistical Surveys. *Math. Geol.* **2007**, *39*, 113–134. [CrossRef]
91. Lark, R.M.; Marchant, B.P. How Should a Spatial-Coverage Sample Design for a Geostatistical Soil Survey Be Supplemented to Support Estimation of Spatial Covariance Parameters? *Geoderma* **2018**, *319*, 89–99. [CrossRef]
92. Wadoux, A.M.J.C.; Marchant, B.P.; Lark, R.M. Efficient Sampling for Geostatistical Surveys. *Eur. J. Soil Sci.* **2019**, *70*, 975–989. [CrossRef]
93. Tóth, T.; Várallyay, G. Past, Present and Future of the Hungarian Classification of Salt-Affected Soils. *Soil Classif.* **2001**, 125–135.
94. Van Orshoven, J.; Terres, J.-M.; Tóth, T. *Updated Common Bio-Physical Criteria to Define Natural Constraints for Agriculture in Europe*; Office for Official Publications of the European Communities: Luxembourg, 2012.
95. Pásztor, L.; Szabó, J.; Bakacsi, Z.; László, P.; Dombos, M. Large-Scale Soil Maps Improved by Digital Soil Mapping and GIS-Based Soil Status Assessment. *Agrokémia és Talajt.* **2007**, *55*, 79–88. [CrossRef]
96. Nabiollahi, K.; Taghizadeh-Mehrjardi, R.; Shahabi, A.; Heung, B.; Amirian-Chakan, A.; Davari, M.; Scholten, T. Assessing Agricultural Salt-Affected Land Using Digital Soil Mapping and Hybridized Random Forests. *Geoderma* **2021**, *385*, 114858. [CrossRef]
97. Kvoda, V.A.; van den Berg, C.; Hagan, R.M. *Irrigation, Drainage and Salinity*; Hutchinson/FAO/UNESCO, 1973. Available online: <https://unesdoc.unesco.org/ark:/48223/pf0000005702> (accessed on 1 June 2022).
98. Buzási, A.; Pálvölgyi, T.; Esses, D. Drought-Related Vulnerability and Its Policy Implications in Hungary. *Mitig. Adapt. Strateg. Glob. Chang.* **2021**, *26*, 1–20. [CrossRef]
99. Sparks, D.L. *Environmental Soil Chemistry*; Academic Press: Cambridge, MA, USA, 1995; ISBN 9780126564457.