Soil salinity mapping using Landsat 8 OLI data and regression modeling in the Great Hungarian Plain

Ghada Sahbeni[1]

Abstract

Salt's deposition in the subsoil is known as salinization. It is caused by natural processes such as mineral weathering or human-made activities such as irrigation with saline water. This environmental issue has grown more critical and is frequently occurring in the Hungarian Great Plain, adversely influencing agricultural productivity. This study aims to predict soil salinity in the Great Hungarian Plain, located in the east of Hungary, using Landsat 8 OLI data combined with four state-of-the-art regression models, i.e., Multiple Linear Regression, Partial Least Squares Regression, Ridge Regression, and Feedforward Artificial Neural Network. For this purpose, seventy-six soil samples were collected during a field survey conducted by the Research Institute for Soil Sciences and Agricultural Chemistry between the 15 of September and the 15 of October, 2016. We used the min–max accuracy, the root-mean-square error (RMSE), and the mean squared error (MSE) to evaluate and compare the four models' performance. The results showed that the ridge regression model performed the best in terms of prediction ($MSE_{training} = 0.006$, $MSE_{test} = 0.0007$, $RMSE = 0.081$), with a min–max accuracy equal to 0.75. Hence, the application of regression modeling on spectral indices, principal component analysis, and land surface temperature derived from multispectral data is an efficient method for soil salinity assessment at local scales. The resulting map can provide an overview of salinity levels and evaluate the efficiency of land management strategies in irrigated areas. An increase in sampling density will be recommended to validate this approach on the regional scale.

1 Introduction

Soil salinization is the salts accumulation process in the subsurface that destroys soil composition and water quality, reduces agricultural productivity, and negatively influences economic growth [1]. As one of the most serious land degradation forms, salinization occurs in arid and semiarid zones, where evaporation exceeds precipitation [2]. The Food and Agriculture Organization (FAO) [3] estimated the salt-affected surface area to be 831 million ha in total, including 434 million ha of sodic soils and 397 million ha of saline soils. In Hungary, salinity and sodicity affect one-third of the Great Hungarian Plain soils, and the potential salt-affected soils (SAS) cover one-third of the territory [4]. Drought, climate change, low water resources, and land-use changes can aggravate salinity conditions [5]. Therefore, authorities tend to adopt monitoring strategies to control this process continuously. Control and prevention approaches on the regional scale are time-consuming and require various resources. As a result, the geographic information systems (GIS) and remote sensing techniques opened the window for technology to replace or assist conventional methods. They established easier, less time-consuming, and inexpensive techniques

✉ Ghada Sahbeni, gsahbeni@caesar.elte.hu | [1]Department of Geophysics and Space Science, Eötvös Loránd University, Pázmány Péter stny. 1/A, Budapest 1117, Hungary.

to assess and predict environmental processes such as salinization. Using spectral response extracted directly from sensors or after the application of spectral transformations, i.e., principal component analysis (PCA) [6, 7], tasseled cap transformation [8], and spectral indices [9, 10], have given promising results in terms of prediction accuracy. Many scholars emphasize the importance of spectral signature in remote sensing studies as a core concept in this field [11–14]. Digital soil mapping has been the focus of several research projects operated with different types of remotely sensed data, using statistical or geostatistical methods. El Hafyani et al. [15] demonstrated Landsat 8 OLI image's efficacity to model soil salinity in Morocco's Tafilalet plain. The results yielded a coefficient of determination R^2 ranges from 0.53 to 0.75, and a root-mean-square error (RMSE) ranges from 0.62 to 0.80 dS/m. Hihi et al. [16] found a medium correlation ($R^2 = 0.48$) between electrical conductivity (EC) and spectral indices derived from a Sentinel 2 MSI image using a simple linear regression model. In Uzbekistan, using variance analysis, Ivushkin et al. [17] found a significant correlation between soil salinity and canopy temperature derived from the moderate resolution imaging spectroradiometer (MODIS) data. The Gaussian processes method applied on SAR Sentinel-1 data and advanced machine learning algorithms produced a highly accurate model that explains the relationship between remotely sensed data and EC, with a coefficient of determination R^2 equal to 0.808 [18]. Taghadosi et al. [19] revealed the efficiency of intensity images derived from VV and VH polarization of SAR Sentinel-1 imagery in the discrimination of saline soils. The support vector regression (SVR) technique with radial basis function (RBF) kernel produced the most accurate model, with a coefficient of determination R^2 equals 0.9783 and an RMSE equals 0.3561. In Algeria, Dehni et al. [10] proved the importance of Landsat ETM+ and ALI-EO-1 images in identifying and delineating saline and sodic soils. Zurqani et al. [20] used a time series of Landsat data for 29 years (1972, 1987, and 2001) and field measurements to detect the spatiotemporal variation of soil salinity in Libya. Weng et al. [21] suggested using hyperspectral imagery to produce a higher accuracy. The univariate regression using a new salinity index (SSI) derived from EO-1 Hyperion data yielded a coefficient of determination R^2 equal to 0.873. Sahbeni G. [22] applied a multiple linear regression analysis between salt content and Sentinel 2 MSI data for modeling soil salinity distribution in the Great Hungarian Plain. The final model showed a significant moderate accuracy with a coefficient of determination R^2 equal to 0.51 and an RMSE equal to 0.1942 g/kg. Li Z. et al. [23] analyzed the performance of multiple linear regression (MLR), geographically weighted regression (GWR), and random forest regression (RFR) models in

terms of soil salinity prediction. The findings revealed that remote sensing imagery scanned during the dry season is better for estimating electrical conductivity (EC), with topographic variables and spatial location playing a crucial part in the modeling. However, the random forest regression (RFR) model had a higher prediction accuracy than the GWR model, while the MLR model had the highest error. Al-Ali Z.M. et al. [24] studied eight different physical models for soil salinity mapping in an arid landscape using spectral reflectance measurements and Landsat 8 OLI data. The statistical tests for models based on visible-near-infrared (VNIR) bands showed insignificant fits with a coefficient of determination R^2 of 0.41 and a very high RMSE (≥ 0.65). In contrast, models based on the second-order polynomial equation using the shortwave infrared (SWIR) bands produced better results, with a coefficient of determination R^2 equal to 0.97 and an RMSE equal to 0.13. The study conducted by Zhang X. and Huang B. [25] to predict soil salinity using soil-reflected spectra showed that smoothing methods and spectral transformations influenced the models' estimation accuracies. The PCR based on the median filtering data smoothing method gave the most accurate results, with a coefficient of determination R^2 equal to 0.7206 and an RMSE equal to 0.3929.

The studies mentioned above have explored the potential efficiency of remotely sensed data in estimating salinity when coupled with field measurements. The electromagnetic signal sensitivity to surface soil parameters can be processed to map salinity distribution with great accuracy. Nevertheless, most of these reviewed studies used statistical analysis and satellite imagery to model soil salinity in the Afro-Eurasia region. Only a few studies focused on the Central-Eastern European zone, although Hungary has the largest expanse of naturally salt-affected soils in the continent.

In flat landscapes with a typical continental climate like the Great Hungarian Plain, soluble salts accumulate more quickly. Certain environmental and climatic conditions make the landscape threatened by salinization, mainly when groundwater levels are oscillating. As a result, modeling soil salinity variability and mapping its spatial pattern for the Great Hungarian Palin become an essential task to endorse successful soil reclamation programs that mitigate possible fluctuations in salinity levels. The aims of this study are to; (i) predict salinity using remotely sensed data derived from Landsat 8 OLI sensor and field measurements through evaluating four regression methods; Multiple Linear Regression (MLR), Partial Least Squares Regression (PLSR), Ridge Regression (RR), and Artificial Neural Network (ANN), and (ii) prove the importance of spectral enhancements in soil salinity modeling.

2 Study area

The study area is located in the Great Hungarian Plain (GHP), east of Hungary (Fig. 1). It lies between longitudes 20°22′24″ and 21°41′51″, and between latitudes 47°5′6″ and 47°48′40″, covering an area of 7602.84 km^2. The average elevation is 88.8 m above sea level. Half of the Great Hungarian Plain (GHP) is flat with topographic level differences of 3–4 m. Wind-blown sand on the hills, loess and loess-like sediments above the level of the actual floodplains, and silty clay in flat alluvial areas are the three deposit types that can be found on its surface [26, 27]. The main river is the Tisza. It flows through the plain and collects tributaries from almost the entire lowland area. The study area climate is semiarid, with an average annual temperature of 11 °C [28]. The most precipitation occurs between May and July, while the least precipitation falls between January and March. The annual mean precipitation is less than 500 mm in the Tisza river's low altitude valley [29].

3 Materials and methods

3.1 Remotely sensed data

We downloaded a Landsat 8 OLI image from the United States Geological Survey (USGS). It was acquired on the 08 of August, 2016. The difference between the imagery acquisition day and the field measurements date is due to the scarcity of cloudless accessible data (Cloud Cover < 5%). Table 1, extracted from the metadata file, illustrates the properties of remotely sensed data used.

SRTM GL1 (30 m Ellipsoidal) is a version of the Shuttle Radar Topography Mission dataset that uses elevation values from WGS84 ellipsoidal height instead of the normal
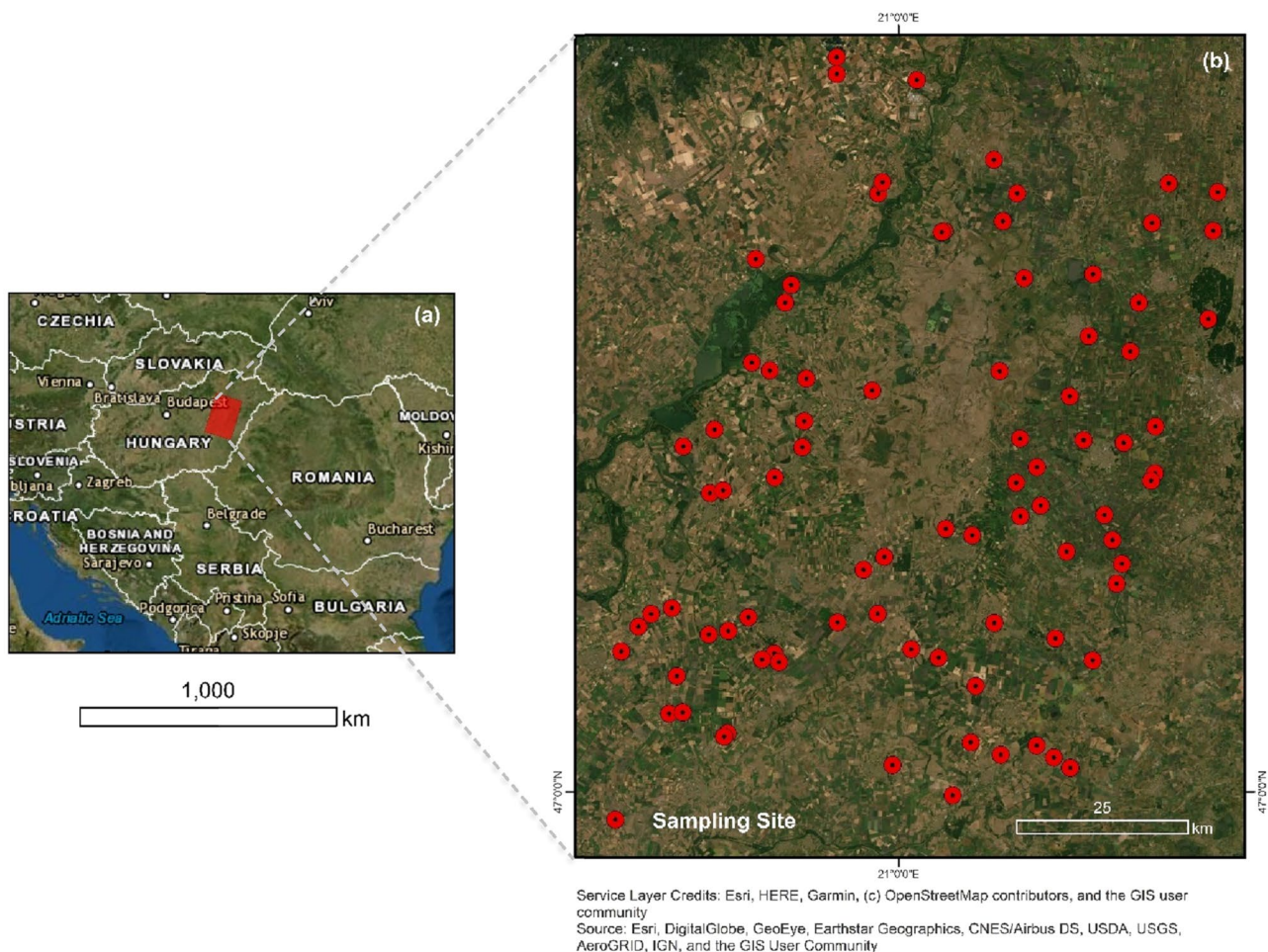

Fig.1 Location of the study area; **a** Satellite imagery map of Hungary (ESRI Basemap, ArcMap 10.3), **b** Sampling sites' locations

Table 1 Properties of Landsat 8 OLI image used

Product ID	LC08_L1TP_187027_20160808_20170322_01_T1
Acquisition date	2016-08-08
Spatial resolution	30 m
Radiometric resolution	8 bits
Output format	GeoTIFF
Sensor	OLI_TIRS
Path/Row	187/27
Map projection	UTM_WGS84_34N

orthometric or geoid-referenced elevation. The EGM96 geoid model was subtracted from the standard SRTM GL1 data to create this dataset. It has a resolution of 1 arc-second (30 m) [30]. We downloaded an SRTM elevation model in GeoTIFF format. The OpenTopography Facility provided it with support from the National Science Foundation [31].

3.2 Soil sampling

The Research Institute for Soil Sciences and Agricultural Chemistry (RISSAC) provided the field measurements database on the 16th of March 2020. Soil samples are collected within the subsoil's upper layer (30 cm) from mid-September until mid-October 2016 [32, 33]. The field campaign is carried out during the dry season to improve salt spectral properties at the surface during salt accumulation. Electrical conductivity is measured in the laboratory by immersing an electrode in a water-saturated soil paste at the plasticity limit [34]. Soil salinity is determined using saturated paste according to the Hungarian Standard MSZ-08-0206/2-1978 [35, 36]. MSZ-08-0206/2-1978 protocol details can be found in (MSZ 1978, 1978) [37].

3.3 Preprocessing

We converted Landsat 8 OLI image to Top of Atmosphere (TOA) Reflectance [54]. Fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) was applied using ENVI 5.3 software package for atmospheric correction with standard options [21, 38, 39].

3.4 Spectral indices

Based on the literature review [40–46], we computed eighteen spectral indices from preprocessed data using ENVI 5.3 and ArcMap 10.3. These indices were selected for this study due to their potential association with salinity detection. Table 2 shows the mathematical expressions of spectral indices.

Table 2 Spectral indices used in regression analyses

Index	Expression
Normalized Difference Vegetation Index NDVI	$(NIR - R)/(NIR + R)$[47]
Normalized Difference Salinity Index NDSI	$(R - NIR)/(R + NIR)$[48]
Vegetation Soil Salinity Index VSSI	$2*G - 5*(R + NIR)$[10]
Albedo	$((0.356*B) + (0.130*R) + (0.373*NIR) + (0.085*SWIR_1)$ [49]
Canopy Response Salinity Index CRSI	$\sqrt{((NIR*R - G*R)/(NIR*R + G*R))}$[5]
Brightness Index BI	$\sqrt{(R^2 + NIR^2)}$[48]
Salinity Index SI	$(R*G)/B$[50]
Salinity Index 1 SI$_1$	$\sqrt{(G*R)}$[44]
Salinity Index 2 SI$_2$	$\sqrt{(R*NIR)}$[10]
Salinity Index 3 SI$_3$	$\sqrt{G^2 + R^2 + NIR^2}$[44]
Salinity Index 4 SI$_4$	$\sqrt{(G^2 + R^2)}$[51]
Salinity Index 5 SI$_5$	B/R[52]
Ratio Spectral Index RSI	R/NIR[53]
Differential Vegetation Index DVI	NIR—R [54]
Intensity Index 1 Int$_1$	$(G + R)/2$[55]
Intensity Index 2 Int$_2$	$(G + R + NIR)/2$[55]
Simple Ratio SR	$(R - NIR)/(G + NIR)$[10]
Land Surface Temperature LST	* [56]

*For detailed information, see (Sobrino et al. 2004); The land surface temperature (LST) calculation method is too extensive to include in this table

The information in intercorrelated variables is condensed into a few variables called principal components. Principal component analysis (PCA) application removes spectral noise and reduces information redundancy. Therefore, we used PCA to compress information from visible, near-infrared, and SWIR bands [6, 7]. Based on the statistics file produced by ENVI 5.3, the first two components explain approximately 99% of data variance. They contain the majority of information that can be useful for soil salinity modeling (Fig. 2).

The effect of topography on salts movement within the profile is revealed by many scholars, mainly where the water balance is negative [57, 58]. Therefore, we included elevation a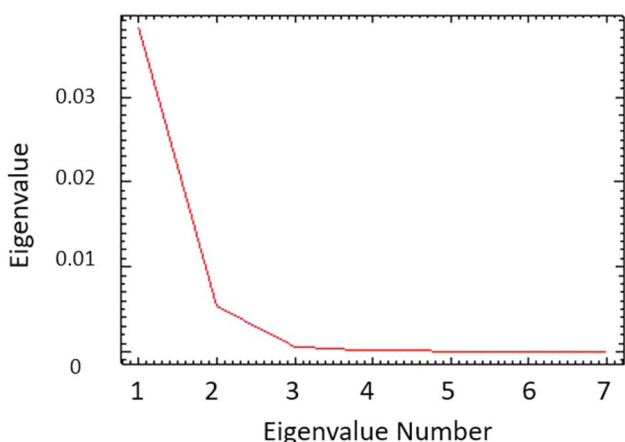s an independent covariate in the regression modeling to evaluate this component's influence on salinity's spatial distribution.

After extracting spectral indices, bands, PCA_1, PCA_2, and elevation values corresponding to sampling sites, we performed regression analyses between salt content measurements and spectral data. The goal is to understand the statistical relationship between remotely sensed data and the sampled soil's salt content. Figure 3 illustrates the study's methodology.

3.5 Regression modeling

On systems with limited computing resources, linear regression is a simple algorithm that can be trained quickly and efficiently. It has a significantly lower time complexity than other machine learning algorithms [59–61].

Due to their robustness to collinearity, which is frequent between multiple variables, ridge regression and partial least squares regression approaches are commonly used [62, 63]. The artificial neural network is distinguished by the efficiency of handling highly complex data, along with the potential to generalize, as discussed by many scholars [46, 64, 65]. This study evaluates the statistical methods cited above. Details are explained and well-presented by *Bishop C.M.* [66].

Multiple linear regression is a linear approach to modeling the relationship between one or more explanatory variables and a scalar variable [16, 50]. Equation (1) describes the relationship between the dependent and explanatory variables.

Fig.2 Eigenvalues and variance accounted by principal components

$$Y = A_0 + A_1 * x_1 + A_2 * x_2 + A_3 * x_3 + \cdots A_n * x_n \qquad (1)$$
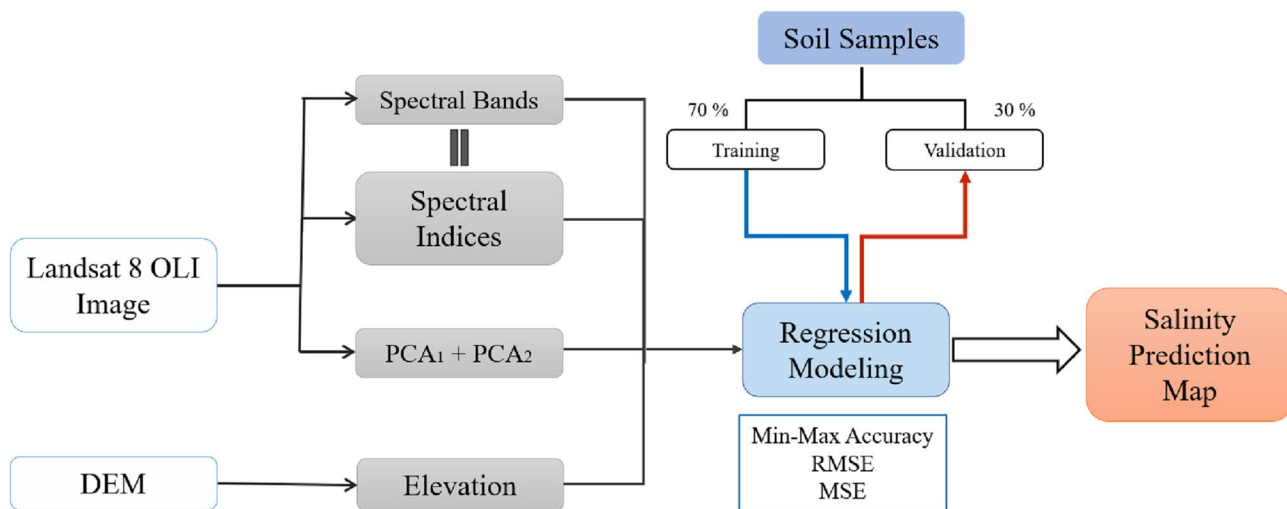

Fig. 3 Schematic diagram of the study methodology

where Y is the dependent variable, x_i is the explanatory variable, A_i is the coefficient of the variable i, and A_0 is the intercept.

By adding a bias to the regression estimates, the ridge regression overcomes the data collinearity problem. The penalization factor or lambda regularizes the coefficients to penalize the optimization function if the coefficients take large values [62, 67]. The ridge estimation of coefficients in a linear regression model $y = \beta(k)*x + b$ is described in Equation (2).

$$\hat{\beta}(k) = (X\prime X + kI) - 1X\prime Y \qquad (2)$$

where X is an n-by-p independent variable matrix, $X\prime$ is its transpose, I is an n-by-n identity matrix, Y is an n-by-1 vector of observations, and k is the ridge parameter.

The Hoel–Kennad formula in Equation (3) can be used to calculate the ridge parameter.

$$k = \hat{\sigma}^2 / \max(\hat{\alpha}\,i2) \qquad (3)$$

where $\hat{\sigma}^2$ is the estimation of residual squares, and $\acute{\sigma}$ is defined as $\acute{\sigma} = \Lambda^{-1}G\prime X\prime Y$, where G is an orthogonal square matrix, and Λ is a diagonal matrix. Λ and G are associated with this equation: $G\prime(X\prime X)G = \Lambda$. [62]

The partial least squares regression (PLSR) reduces the predictors to a smaller collection of uncorrelated components before performing least squares regression on these components. This approach is advantageous when the predictors are strongly collinear. The predictors are measured with error, making PLSR more robust to uncertainty measurement [63, 68].

$$y = \sum_{1}^{n} w_i * \rho_i + w_0 \qquad (4)$$

where y is the dependent variable, ρ is the explanatory variable, w is the model coefficient for the ith variable, w_0 is the constant item, and n is the number of variables.

The artificial neural network (ANN) is a problem-solving model inspired by the biological neuron system and consisted of many highly interconnected computing elements called neurons. It takes a nonlinear path and processes data in parallel through nodes, resulting in a complex adaptive system that can alter its internal structure by changing input weights [69]. In this study, we used a feedforward neural network with a single hidden layer. Its simplicity distinguishes it from other types of artificial neural networks, in which data moves through various input nodes before it reaches the output node [46, 70]. For a regression function $y = f(x)$, a feedforward network defines a mapping $y = f(x; \theta)$ and learns the value of the parameters θ, resulting in the best function approximation, described in Eq. (5).

$$y = f\left(\sum_{1}^{n} x_i * w_i + b \right) \qquad (5)$$

where y is the dependent variable, x is the explanatory variable, n is the number of independent variables (i goes from 1 to n), w is the weight of the explanatory variable, and b is the bias added to the function. Figure 4 illustrates the neural network's basic concept.

We used seventy-six soil samples in this study, with 70% for training and 30% for validation. Three statistical metrics, i.e., root-mean-square error (RMSE), mean squared error (MSE), and min-max accuracy, are computed to assess regression models' performance.

Equations (6) and (7) describe the mathematical expressions of MSE and RMSE.

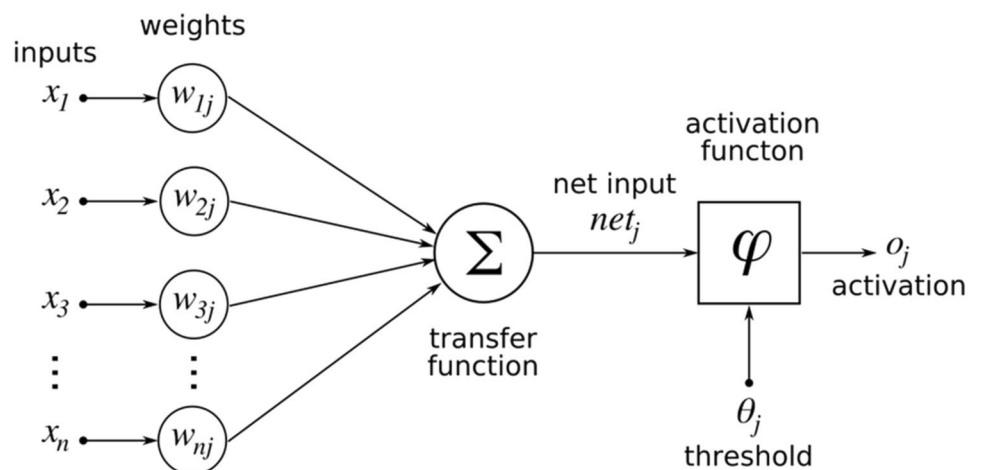**Fig. 4** Basic concept of artificial neural network [71]

Table 3 Descriptive statistics on salt content samples

	Maximum	Minimum	Mean	Median	Skewness
Salt content (%)	0.26	0	0.057	0.06	1.96

Table 4 Models' performance using three statistical parameters

Model	MSE		RMSE	Accuracy
	Training	Test		
MLR	0.004	0.014	0.004	0.73
PLSR	0.006	0.009	0.006	0.73
RR	0.006	0.007	0.081	0.75
ANN	0.029	0.108	0.118	0.71

$$RMSE = \sqrt{\sum_{1}^{n} (\hat{y}_i - y_i)/n} \qquad (6)$$

$$MSE = \frac{1}{n} * \sum_{1}^{n} (y_i - \hat{y}_i) \qquad (7)$$

where $\hat{y}_i$ is the predicted value for the ith observation, y_i is the measured value for the ith observation, and n is the total number of observations.

4 Results and discussion

4.1 Descriptive analysis

The main statistical parameters for the salt content data are given in Table 3. The disparity between the minimum and the maximum reflects a spatial variability in salinity levels. The normality test revealed that the salt content data were positively biased. The contrast between the mean and the median has demonstrated this point.

We applied a square root transformation to increase the distribution of data to normality. This was approved by a remarkable improvement in skewness from 1.96 for the initial data to -0.39 for the normalized data.

Fig. 5 Relationship between measured and predicted normalized salt content values, **a** MLR, **b** PLSR, **c** RR, **d** ANN
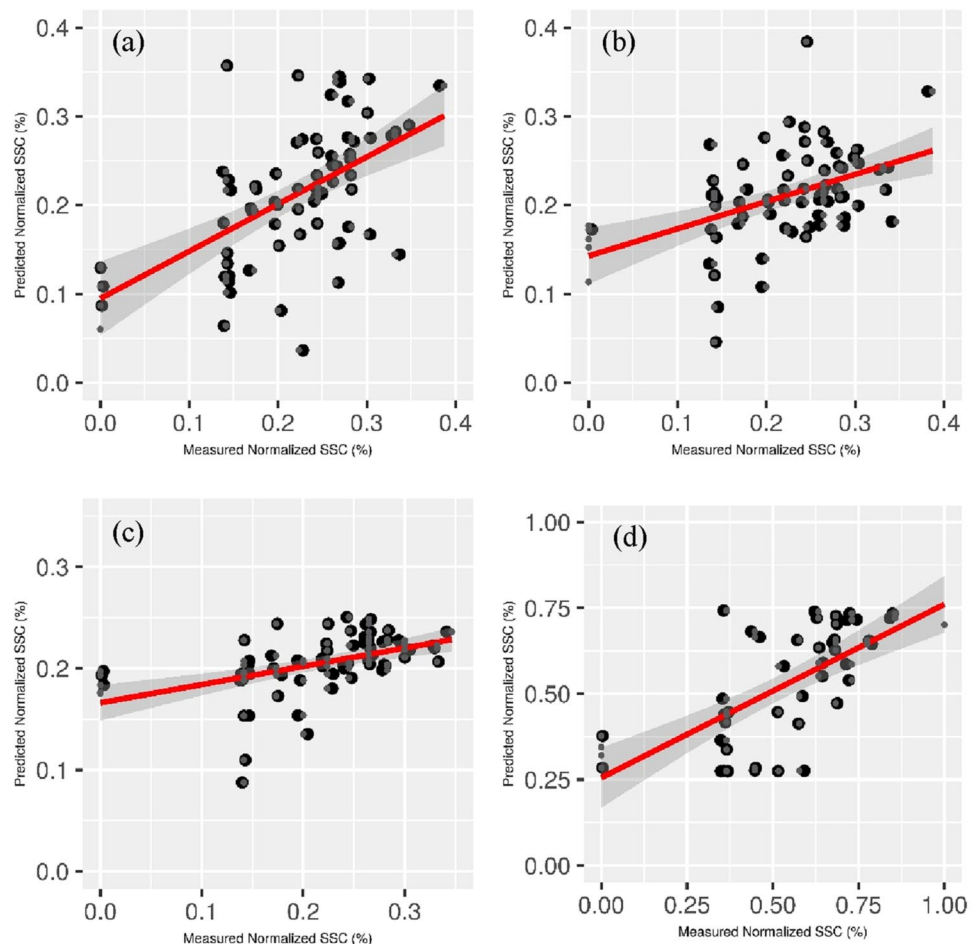
4.2 Models performance

A low RMSE value indicates an accurate prediction. Similarly, a low MSE value indicates a more precise estimation. We calculated the statistical parameters using Eqs. (6) and (7). More RMSE and MSE are closer to zero, higher is the model's accuracy. The statistical performance of the regression models is given in Table 4. Overall, the four regression models have a satisfactory goodness-of-fit for the training and test sets, based on Table 4 and Fig. 5. The MSE of the ridge regression (RR) model is the lowest for the test sample (= 0.007). The MLR model has the lowest MSE (= 0.004) for the training sample, followed by the RR and the PLSR models (MSE=0.006), and eventually the ANN model (MSE=0.029). The MLR model yielded the lowest RMSE value (= 0.004), followed by the PLSR model (= 0.006), the RR model (= 0.081), and finally, the ANN model (= 0.118). To evaluate the models' accuracy, we ran a min-max accuracy test, which takes the mean values of the minimum and maximum values between the measured and the predicted values. It has a range from zero to one. A min-max accuracy score equal to one indicates a perfect match between the actual and estimated values [72]. The RR model has the highest accuracy (= 0.75), followed by the PLSR model (= 0.73), the MLR (= 0.73), and the feedforward ANN model (= 0.71).

The objectives to be achieved determines the appropriate approach for a model's selection. Overall, the highest accuracy model would be favored, but very low RMSE and MSE must also be considered. Hence, the ridge regression (RR) model is regarded as the best model with the highest accuracy (= 75 %) and the lowest MSE for the test set.

4.3 Salinity prediction map

Based on three statistical metrics analyses, the ridge regression model was selected for predicting soil salinity in the Great Hungarian Plain. Therefore, we used Eq. (8).

The maximum predicted value of normalized salt content equals 0.38. The minimum estimated value equals -0.13. The mean value equals 0.21, and the standard deviation equals 0.03. Water bodies and urban infrastructures are the most common negative predictions. As a result, we used a masking tool to exclude these pixels from the landscape. According to Weng et al. [21], negative predictions are illogical and indicative of prediction errors. However, it can also be caused by residual noise that occurs after atmospheric corrections.

As data rescaling represents an essential step in data analysis, we converted the salt content variable from (%) to g/kg, giving it broader ranges for the classification (Table 5). The final map is presented in Fig. 6.

Figure 6 shows that around 99% of total classified pixels are non-saline and less than 1% are low saline. The maximum predicted value of salt content equals 1.46 g/kg. In contrast, the minimum estimated value equals zero. The mean equals 0.44 g/kg, and the standard deviation equals 0.11 g/kg.
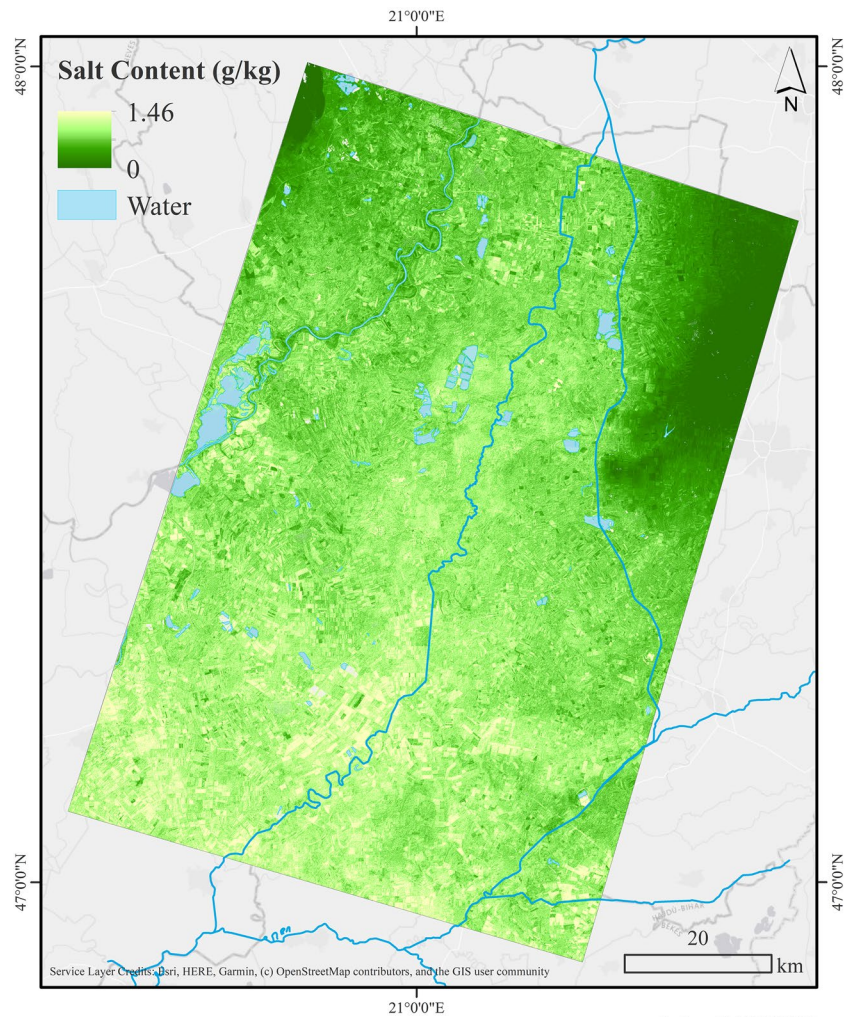
A visual interpretation of the map showed a noticeable association between extremely non-saline soils and the elevation. Despite the study area's absolute flatness, the effect of microrelief in salinity distribution is relatively evident on the map. Comparing LST variation with soil salinity distribution, we found similarities between the two variables' spatial distribution. High LST values are frequent in soils with higher salinity, while lower LST values occupy firmly non-saline soils (0 g/kg). The correlation matrix derived from the band collection statistics test shows a strong negative correlation with elevation, and a moderately low positive correlation with LST, with a coefficient of determination R^2 equal to -0.86 and 0.28, respectively.

Overall, saline land features discrimination is an achievable task using spectral response values and spectral transformations derived from multispectral sensors. This agrees with Matinfar et al. [73] and Wu et al. [74] results regarding spectral indices' efficiency in

$$
\begin{aligned}
SQRT(SSC\%) = {}& 1.36582 + 0.00087 * LST - 0.03777 * CRSI - 0.00103 * BI + 0.1195 * SI - \\
& 0.01713 * SI_1 - 0.05036 * SI_2 + 0.016 * SR + 0.0141 * NDSI - 0.25137 * \log(elevation) + \\
& 0.00059 * VSSI + 0.02754 * RSI - 0.1545 * PCA_1 + 0.03208 * PCA_2 + 0.00106 * SI_3 - \\
& 0.00184 * SI_5 - 0.0083 * DVI - 0.10293 * B_7
\end{aligned} \tag{8}
$$

Table 5 Soil salinity content levels using Chinese Saline Soil Classification Standard presented by WENG et al. [21]

SSC*	< 1 g/kg	1–2 g/kg	2–4 g/kg	4–6 g/kg	> 6 g/kg
Class	Non-saline	Low saline	Moderately saline	Highly saline	Extremely saline

*SSC: Soil Salinity Content

Fig.6 Soil salinity prediction map using the final model

Author: G. SAHBENI
Data Source: USGS (Landsat 8 OLI image), OpenTopography (DTM) , RISSAC (Soil samples)

soil salinity mapping. Mousavi et al. [39] found a strong association between spectral indices and soil salinity distribution using artificial neural network ($R^2 = 0.964$), which is relatively in agreement with our findings. Nevertheless, the ANN model performed the least among other regression models producing the lowest accuracy (= 71%) and the highest RMSE (RMSE= 0.118). The small sample size used to train the neural network limited the prediction performance of this method. Using a multiple linear regression model, Hihi et al. [16] found a medium correlation between spectral indices derived from Sentinel 2 MSI sensor and salt content in the Tunisian South, with a determination coefficient R^2 equals 0.48. Allbed et al. [50] found a medium correlation with a coefficient of determination R^2 equals 0.65 using an IKONOS image and regression techniques in Al-Hassa oasis, Saudi Arabia. Asfaw et al. [2] found a strong relationship between remotely sensed data and electrical conductivity (EC) in a regression model with a coefficient of determination

R^2 equals 0.78. These findings agree with our approach, elucidating statistical modeling's potential to map soil salinity with lower costs.

Furthermore, this work demonstrates the significance of shortwave infrared bands, and in particular the second SWIR band, in salinity prediction. This agrees with Hihi S et al. [16] and Lamqadem A et al. [75] studies. Bannari et al. [76] also revealed the importance of SWIR bands in soil salinity modeling using Sentinel 2 MSI image. Electrical conductivity in the laboratory (EC_{Lab}) showed a moderate association with SWIR bands with a coefficient of determination R^2 equal to 0.5 for $SWIR_1$ and 0.6 for $SWIR_2$. The relationship is uncertain for sand patterns, moist soils, and vegetated areas with high salinity values and low reflectance in the near-infrared channel, where prediction errors are more likely to occur. Future studies will focus on using data derived from hyperspectral sensors to overcome this issue, as recommended by previous studies [21, 53, 77].

Despite models' accuracy differences from one study area to another, a significant relationship between remotely sensed data and soil salinity has been well-established in the literature. Moreover, several environmental factors influence soil salinity measurements and spectral information reception, explaining the findings disparity within previous studies.

Like many natural processes, soil salinization is influenced by environmental variations, making it difficult to track and analyze with absolute certainty. The prediction errors persist even with using the most robust statistical technique. Therefore, it is suggested to perform a validation sampling to evaluate the regression model's applicability and estimate its real predictive potential.

5 Conclusions

Soil salinization is a major environmental threat that affects soil and water quality, reduces agricultural production and food safety worldwide. As a result, collecting valuable information on salinity levels is a critical challenge for decision makers to improve and sustain land management strategies.

We discussed this issue by evaluating the usefulness of Landsat 8 OLI imagery for soil salinity prediction using four state-of-the-art regression modeling algorithms. Accordingly, the ridge regression approach performed the best in this research with a penalty parameter equal to 0.387 and a min-max accuracy equal to 75%. The model can be applied in another study area with similar climatic conditions and geological formations. As revealed by many scholars in the literature section, Landsat 8 OLI image has shown its relative efficiency in soil salinity estimation. This study demonstrates the importance of spectral enhancements in mapping environmental parameters such as soil salinity and identifying pattern variations. Hence, the elevation is proved to have a statistical significance for soil salinity prediction, with a p-value equal to 0.002. Throughout the combination of field measurements and spectral information, remote sensing represents an inexpensive and effective reliable tool to estimate soil salinity.

Future studies will focus on reducing spectral noise, improving the model's predictive power, and investigating the role of environmental factors in salts distribution and movement.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

References

1. Lhissoui R, Harti AE, Chokmani K (2014) Mapping soil salinity in irrigated land using optical remote sensing data. Eurasian J Soil Sci EJSS 3(2):82. https://doi.org/10.18393/ejss.84540
2. Asfaw E, Suryabhagavan KV, Argaw M (2018) Soil salinity modeling and mapping using remote sensing and GIS: the case of wonji sugar cane irrigation farm, Ethiopia. J Saudi Soc Agric Sci 17(3):250–258. https://doi.org/10.1016/j.jssas.2016.05.003
3. FAO (2000) Extend and causes of salt-affected soils in participating countries, Global Network on Integrated Soil Management for Sustainable Use of Salt-affected soils, FAO-AGL website
4. Tóth T (1991) Monitoring, predicting, and quantifying soil salinity, sodicity, and alkalinity in Hungary at a different scale. Past experiences, current achievements, and an outlook with special regard to the European Union
5. Scudiero E, Corwin DL, Anderson RG, Yemoto K, Clary W, Wang Z, Skaggs TH (2017) Remote sensing is a viable tool for mapping soil salinity in agricultural lands. Calif Agric 71(4):231–238. https://doi.org/10.3733/ca.2017a0009
6. Ten Caten A, Dalmolin RSD, de Araújo Pedron F, Santos MDLM (2011) Principal components as predictor variables in digital mapping of soil classes. Cienc. Rural [online]. 41(7), 1170-1176. ISSN 0103-8478. Doi: https://doi.org/10.1590/S0103-84782011000700011
7. Sahu SK, Prasad MBNV, Tripathy BK (2015) PCA Classification technique of remote sensing analysis of colour composite image of chillika lagoon, Odisha. Int J Adv Res Comput Sci Softw Eng 5(5):513–518
8. Li B, Ti C, Zhao Y, Yan X (2016) Estimating soil moisture with Landsat data and its application in extracting the spatial distribution of winter flooded paddies. Remote Sens. https://doi.org/10.3390/rs8010038

9. Wu W (2014) The generalized difference vegetation index (gdvi) for dryland characterization. Remote Sens 6(2):1211–1233. https://doi.org/10.3390/rs6021211

10. Dehni A, Lounis M (2012) Remotesensingtechniquesforsalt-affectedsoilmapping: application to the Oran region of Algeria. Proced Eng 33:188–198. https://doi.org/10.1016/j.proeng.2012.01.1193

11. Van Leeuwen W (2009) Visible near IR and shortwave IR spectral characteristics of terrestrial surfaces. In: The SAGE handbook of remote sensing, SAGE Publications Inc., pp 32–50. Doi: https://doi.org/10.4135/9780857021052

12. Aceves EÁ, Guevara HJP, Enríquez AC, Gaxiola JDJC, Cervantes MDJP, Barrientos JH, Samuel CL (2019) Determiningsalinity-andionsoilusingsatelliteimageprocessing. Pol J Environ Stud 28(3):1549–1560. https://doi.org/10.15244/pjoes/81693

13. Ling C, Liu H, Ju H, Zhang H, You J, Li W (2017) A study on spectral signature analysis of wetland vegetation based on ground imaging spectrum data. Journal of Physics: Conference Series: The 2017 International Conference on Cloud Technology and Communication Engineering (CTCE2017), 910. pp 18–20 August 2017, Guilin, China

14. Ji W, Adamchuk V, Chen S, Biswas A, Leclerc M, et al. (2017) The use of proximal soil sensor data fusion and digital soil mapping for precision agriculture. Pedometrics, Jun 2017, Wageningen, Netherlands. 298. ffhal-01601278f

15. El hafyani M, Essahlaoui A, El Baghdadi M et al (2019) Modeling and mapping of soil salinity in tafilalet plain (Morocco). Arab J Geosci 12:35. https://doi.org/10.1007/s12517-018-4202-2

16. Hihi S, Rabah ZB, Bouaziz M, Chtourou MY, Bouaziz S (2019) Prediction of soil salinity using remote sensing tools and linear regression model. Adv Remote Sens 08(03):77–88. https://doi.org/10.4236/ars.2019.83005

17. Ivushkin K, Bartholomeus H, Bregt AK, Pulatov A (2017) Satellite thermography for soil salinity assessment of cropped areas in Uzbekistan. Land Degrad Dev 28(3):870–877. https://doi.org/10.1002/ldr.2670

18. Hoa PV, Giang NV, Binh NA, Hai LVH, Pham TD, Hasanlou M, Bui DT (2019) SoilsalinitymappingusingSARSentinel-1dataandadvancedmachinelearningalgorithms: a case study at Ben Tre province of the Mekong River Delta (Vietnam). Remote Sens. https://doi.org/10.3390/rs11020128

19. Taghadosi M, Hasanlou M, Eftekhari K (2018) Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. Int J Remote Sens 2018:1–16

20. Zurqani H, Nwer B, Rhoma E (2012) Assessment of spatial and temporal variations of soil salinity using remote sensing and geographic information system in Libya. Global Science & Technology Forum Pte Ltd. https://doi.org/10.5176/2251-3361_geos12.64

21. Weng YL, Gong P, Zhu ZL (2010) A spectral index for estimating soil salinity in the yellow river delta region of China using EO-1 hyperion data. Pedosphere 20(3):378–388. https://doi.org/10.1016/S1002-0160(10)60027-6

22. Sahbeni G (2020) A linear regression model for soil salinity prediction in The Great Hungarian Plain using Sentinel 2 Data. COPERNICUS 4 BULGARIA 2020: II-nd National Workshop with international participation under the EU Copernicus Programme. Doi: https://doi.org/10.13140/RG.2.2.33536.56320

23. Li Z, Li Y, Xing A et al (2019) Spatial prediction of soil salinity in a Semiarid Oasis: environmental sensitive variable selection and model comparison. Chin Geogr Sci 29:784–797. https://doi.org/10.1007/s11769-019-1071-x

24. Al-Ali ZM, Bannari A, El-Battay RH, ., Shahid, S.A., Hameid, N. A (2021) Validation and comparison of physical models for soil salinity mapping over an arid landscape using spectral reflectance measurements and Landsat-OLI data. Remote Sens 2021(13):494. https://doi.org/10.3390/rs13030494

25. Zhang X, Huang B (2019) Prediction of soil salinity with soil-reflected spectra: a comparison of two regression methods. Sci Rep 9:5067. https://doi.org/10.1038/s41598-019-41470-0

26. Kiss T, Hernesz P, Sümeghy B, Györgyövics K, Sipos G (2014) The evolution of the Great Hungarian Plain fluvial system e fluvial processes in a subsiding area from the beginning of the Weichselian. Quat Int. https://doi.org/10.1016/j.quaint.2014.05.050

27. A Ronai (1986) The quaternary of the great Hungarian Plain English summary Loess Letter, Supplement 12

28. Tóth T, Balog K, Szabo A, Pásztor L, Jobbágy EG, Nosetto MD, Gribovszki Z (2014) Special issue: physiology and ecology of halophytes-plants living in salt-rich environments: Influence of lowland forests on subsurface salt accumulation in shallow groundwater areas. AoB Plants. https://doi.org/10.1093/aobpla/plu054

29. Hungarian Meteorological Service (2018) Precipitation conditions of Hungary. Retrieved from:https://www.met.hu/en/eghajlat/magyarorszag_eghajlata/altalanos_eghajlati_jellemzes/csapadek/ Accessed: 2021–03–12

30. Open Topography (2017) Three new global topographic datasets available (SRTM Ellipsoidal, ALOS World 3D, GMRT). Retrieved from: https://opentopography.org/news/three-new-global-topographic-datasets-available-srtm-ellipsoidal-alos-world-3d-gmrt Accessed: 2021–03–12

31. NASA Shuttle Radar Topography Mission (SRTM) (2013) Shuttle radar topography mission (SRTM) global. Distrib OpenTopogr. https://doi.org/10.5069/G9445JDFAccessed:2021-03-12

32. J Szabó, B Pirkó (2017) Sciences, S., & Chemistry, A. (n.d.). The soil information and monitoring system (TIM)

33. Judit B-Ü, Dóra S, Talajv M, Rendszer M (2016). A talajmonitoring rendszer hungarian soil monitoring system Soil Monitoring System (TIM)

34. Kádár I (1998) 2 MANUAL remediation of contaminated soils assessment, Responsible publisher: Ministry of the Environment. Available from: http://fava.hu/kvvm/www.kvvm.hu/szakmai/karmentes/kiadvanyok/karmkezikk2/2-09.htm

35. Pásztor L (2021) Details about the protocol of soil sampling and salinity measurement, email message to Ghada Sahbeni, the 01st of January

36. European commission joint research centre institute for environment and sustainability (2013) European Hydro-pedological Data Inventory (EU-HYDI)

37. MSZ 1978 (1978) Determination of total water-soluble salt content. (Vízben oldható összes sótartalom meghatározása). Hungarian Standard no. MSZ 08–0206–2:1978. Hungarian Standards Institution. Budapest (in Hungarian)

38. Young NE, Anderson RS, Chignell SM, Vorster AG, Lawrence R, Evangelista PH (2017) A survival guide to Landsat preprocessing. Ecology 98(4):920–932. https://doi.org/10.1002/ecy.1730

39. Mousavi SZ, Habibnejad M, Kavian A, Solaimani K, Khormali F (2017) Digitalmapping of topsoil salinity using remote sensing indices in. Ecopersia 5(2):1771–1786. https://doi.org/10.18869/modares.Ecopersia.5.2.1771

40. Gorji T, Yildirim A, Sertel E, Tanik A (2019) Remote sensing approachesandmappingmethodsformonitoringsoilsalinityunderdifferentclimateregimes. Int J Environ Geoinf IJEGEO 6(1):33–49. https://doi.org/10.30897/ijegeo

41. Hernández EI, Melendez-Pastor I, Navarro-Pedreño J, Gómez I (2014) Spectral indices for the detection of salinity effects in melon plants. Scientia Agricola 71(4):324–330. https://doi.org/10.1590/0103-9016-2013-0338

42. Nouri H, Borujeni SC, Alaghmand S, Anderson SJ, Sutton PC, Parvazian S, Beecham S (2018) Soil salinity mapping of urban

greenery using remote sensing and proximal sensing techniques; the case of veale gardens within the Adelaide parklands. Sustain Switzerland. https://doi.org/10.3390/su10082826

43. Azabdaftari A, Sunar F (2016). Soil salinity mapping using multitemporal Landsat data. In: International archives of the photogrammetry, remote sensing and spatial information sciences-ISPRS Archives; International Society for Photogrammetry and Remote Sensing. 41: pp 3–9. 10.5194/isprsarchives-XLI-B7-3-2016

44. Douaoui AEK, Nicolas H, Walter C (2006) Detecting salinity hazards within a semi-arid context by means of combining soil and remote-sensing data. Geoderma 134(1–2):217–230. https://doi.org/10.1016/j.geoderma.2005.10.009

45. Guo B, Han B, Yang F, Fan Y, Jiang L, Chen S, Yang W, Gong R, Liang T (2019) Salinization information extraction model based on VI–SI feature space combinations in the Yellow River Delta based on Landsat 8 OLI image. Geomat Nat Haz Risk 10(1):1863–1878. https://doi.org/10.1080/19475705.2019.1650125

46. Morgan RS, El-Hady MA, Rahim IS (2018) Soil salinity mapping utilizingsentinel-2 and neural networks. Indian J Agric Res 52(5):524–529. https://doi.org/10.18805/IJARe.A-316

47. Rouse JW, Haas RH, Schell JA, Deering DW (1974) Monitoring vegetation systems in the Great Plains with ERTS. In: Freden SC, Mercanti EP, Becker M (eds) Third earth resources technology satellite–1 symposium, vol I. Technical Presentations, NASA SP-351. NASA, Washington, pp 309–317

48. Khan NM, Rastoskuev VV, Sato Y, Shiozawa S (2005) Assessment of hydro saline land degradation by using a simple approach of remote sensing indicators. Agric Water Manage 77:96–109

49. Silva BB. da, Braga AC, Braga CC, Oliveira de LMM, Montenegro SMGL, Barbosa Junior B (2016) Procedures for calculation of the albedo with OLI-Landsat 8 images: application to the Brazilian semi-arid. Revista Brasileira de Engenharia Agrícola e Ambiental, 20(1): 3–8. https://doi.org/10.1590/1807-1929/agriambi.v20n1p3-8

50. Allbed A, Kumar L, Sinha P (2014) Mapping and modeling spatial variation in soil salinity in the Al Hassa oasis based on remote sensing indicators and regression techniques. Remote Sens 6(2):1137–1157. https://doi.org/10.3390/rs6021137

51. Yahiaoui I, Douaoui A, Zhang Q, Ziane A (2015) Soil salinity prediction in the lower cheliff plain (Algeria) based on remote sensing andtopographicfeature analysis. J Arid Land. https://doi.org/10.1007/s40333-015-0053-9

52. Abbas A, Khan S. (2007). Using remote sensing techniques for appraisal of irrigated soil salinity. In: International Congress on Modelling and Simulation of Australia and New Zealand Christchurch, New Zealand. pp 2632–8

53. Kahaer Y, Tashpolat N (2019) Estimating salt concentrations based on optimized spectral indices in soils with regional heterogeneity. J Spectrosc. https://doi.org/10.1155/2019/2402749

54. Basso F, Bove E, Dumontet S, Ferrara A, Pisante M, Quaranta G, Taberner M (2000) Evaluating environmental sensitivity at the basin scalethroughtheuseofgeographicinformationsystemsandremotelysenseddata:anexamplecoveringtheAgribasin(SouthernItaly). Catena 40:19–35. https://doi.org/10.1016/S0341-8162(99)00062-4

55. Bouaziz M, Matschullat J, Gloaguen R (2011) Improved remote sensing detection of soil salinity from a semi-arid climate in Northeast Brazil. CR Geosci 343:795–803

56. Sobrino JA, Jiménez-Muñoz JC, Paolini L (2004) Land surface temperature retrieval from LANDSAT TM 5. Remote Sens Environ 90(4):434–440. https://doi.org/10.1016/j.rse.2004.02.003

57. Zhao Y, Feng Q, Yang H (2016) Soil salinity distribution and its relationship with soil particle size in the lower reaches of Heihe River Northwestern China. Environ Earth Sci 75:810. https://doi.org/10.1007/s12665-016-5603-8

58. Cramer V, Hobbs R, Atkins L (2004) The influence of local elevation on the effects of secondary salinity in remnant eucalypt woodlands: Changes in understorey communities. Plant and Soil. 265(1/2): 253–266. Retrieved the 15th of March, 2021, from http://www.jstor.org/stable/42951599

59. Jung C, Lee Y, Cho Y, Kim S (2017) A study of spatial soil moisture estimationusing a multiple linear regression model and MODIS land surface temperature data correctedbyconditional merging. Remote Sens. https://doi.org/10.3390/rs9080870

60. Wicki A, Parlow E (2017) Multiple regression analysis for unmixing of surfacetemperaturedatainanurbanenvironment. Remote Sens. https://doi.org/10.3390/rs9070684

61. Mansouri E, Feizi F, Jafari Rad A, Arian M (2018) Remote-sensing data processing with the multivariate regression analysis method for iron mineral resource potential mapping: a case study in the Sarvian area, central Iran. Solid Earth 9:373–384. https://doi.org/10.5194/se-9-373-2018

62. Cai T, Cunyong Ju, Yang X (2009) Comparison of ridge regression and partial least squares regression for estimating aboveground biomass with Landsat images and Terrain Data in Mu Us Sandy Land China. Arid Land Res Manag 23(3):248–261. https://doi.org/10.1080/15324980903038701

63. Cho MA, Skidmore A, Corsi F, van Wieren SE, Sobhan I (2007) Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. Int J Appl Earth Observ Geoinf 9:414–424

64. Arruda GP de, Demattê JAM, Chagas da CS, Fiorio PR, Souza e AB, Fongaro CT (2016) Digital soil mapping using reference area and artificial neural networks. Scientia Agricola 73(3): 266–273 https://doi.org/10.1590/0103-9016-2015-0131

65. Jiang H, Rusuli Y, Amuti T, He Q (2018) Quantitative assessment of soil salinity using multi-source remote sensing data based on the support vector machine and artificial neural network. Int J Remote Sens 2018:1–23

66. Bishop CM (2006) Pattern recognition and machine learning information science and statistics. Springern, Secaucus

67. Hernandez J, Lobos GA, Matus I, del Pozo A, Silva P (2015) Galleguillos M (2015) Using ridge regression models to estimate grain yield from field spectral data in bread wheat (Triticum Aestivum L.) grown under three water regimes. Remote Sens 7:2109–2126. https://doi.org/10.3390/rs70202109

68. Fan X, Liu Y, Tao J, Weng Y (2015) Soil salinity retrieval from advanced multispectral sensor with partial least square regression. Remote Sens 7(1):488–511. https://doi.org/10.3390/rs70100488

69. Navlani A (2019) Neural network models in R. Data camp. Retrieved from: https://www.datacamp.com/community/tutorials/ neural-network-models-r

70. Nevtipilova V, Pastwa J, Boori MS, Vozenilek V (2014) Testing artificial neural network (ANN) for spatial interpolation. J Geol Geosci 3:145. https://doi.org/10.4172/2329-6755.1000145

71. Dickson B (2019) What are artificial neural networks (ANN)? Retrieved from: https://bdtechtalks.com/2019/08/05/what-is-artificial-neural-network-ann/

72. Zhang X-Y, Huang Z, Su X, Siu A, Song Y, Zhang D et al (2020) Machine learning models for net photosynthetic rate prediction using poplar leaf phenotype data. PLoS ONE 15(2):e0228645. https://doi.org/10.1371/journal.pone.0228645

73. Matinfar HR, Zandieh V (2016) Efficiency of spectral indices derived from Landsat-8 images of Maharloo Lake and its surrounding rangelands. J Rangeland Sci 6(4):334–343

74. Wu W, Al-Shafie WM, Mhaimeed AS, Ziadat F, Nangia V, Payne WB (2014) Soil salinity mapping by multiscale remote sensing in Mesopotamia, Iraq. IEEE J Sel Top Appl Earth Obs Remote Sens 7(11):4442–4452. https://doi.org/10.1109/JSTARS.2014.2360411

75. Lamqadem A, Saber H, Rahimi A (2018) Mapping soil salinity using Sentinel-2 image in Ktaoua oasis (Southeast of Morocco), 7th Digital Earth Summit 2018, APRIL 17-19. EL Jadida, Morocco

76. Bannari A, El-Battay A, Bannari R, Rhinane H (2018) Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape. Remote Sens 10:855

77. G Metternicht, Zinck AJ (2008) Soil salinity and salinization hazard. In: Remote sensing of soil salinization. CRC Press, Boca Raton. https://doi.org/10.1201/9781420065039.pt1