Research Article

Ghada Sahbeni*

A PLSR model to predict soil salinity using Sentinel-2 MSI data

<https://doi.org/10.1515/geo-2020-0286>

received September 24, 2020; accepted July 22, 2021

Abstract: Salinization is one of the most widespread environmental threats in arid and semi-arid regions that occur either naturally or artificially within the soil. When exceeding the thresholds, salinity becomes a severe danger, damaging agricultural production, water and soil quality, biodiversity, and infrastructures. This study used spectral indices, including salinity and vegetation indices, Sentinel-2 MSI original bands, and DEM, to model soil salinity in the Great Hungarian Plain. Eighty-one soil samples in the upper 30 cm of the soil surface were collected from vegetated and nonvegetated areas by the Research Institute for Soil Sciences and Agricultural Chemistry (RISSAC). The sampling campaign of salinity monitoring was performed in the dry season to enhance salt spectral characteristics during its accumulation in the subsoil. Hence, applying a partial least squares regression (PLSR) between salt content (g/kg) and remotely sensed data manifested a highly moderate correlation with a coefficient of determination R^2 of 0.68, a p -value of 0.000017, and a root mean square error of 0.22. The final model can be deployed to highlight soil salinity levels in the study area and assist in understanding the efficacy of land management strategies.

Keywords: soil salinity, Sentinel-2 MSI, PLSR, regression analysis, multispectral remote sensing, statistical modeling, the Great Hungarian Plain

1 Introduction

Soil salinization is a severe land degradation form in arid and semi-arid areas where evaporation exceeds precipitation [1]. The salt-affected surface covers 831 million ha, including 434 million ha of sodic soils and 397 million ha

of saline soils [2]. One-third of the Great Hungarian Plain soils are affected by salinity and sodicity, and one-third of the territory is covered by potential salt-affected soils (SAS) [3].

Drought, climate change, low water resources, and land-use changes can aggravate salinity conditions, leading to extreme salinization [4]. The dynamic nature of this phenomenon requires regular monitoring to keep up-to-date information about their extent, severity level, and spatial variation. Nevertheless, the magnitude of such an environmental process makes it challenging to model and analyze on a regional scale [5]. While using conventional approaches to monitor soil salinity on local or regional scales is time consuming and requires enormous resources, geographic information systems (GIS) and remote sensing tools have become suitable alternatives, creating easier, less time consuming, and more affordable methods to assess and control environmental threats. Analyzing spectral response to identify and qualify soil characteristics represents the core of digital soil mapping (DSM), as discussed by many scholars [6–9]. Diverse statistical, geostatistical, and machine learning methods are used to model and measure the uncertainty of DSM outputs, where the soil parameter of interest is considered a realization of a random variable at a single location [10]. Original bands, spectral enhancement techniques, e.g., principal component analysis (PCA) [11,12], Tasseled Cap transformation [13], and spectral indices [14], have given relatively acceptable results with regards to mapping soil parameters.

In the last decades, soil salinity mapping has become a central topic for many studies using multispectral, hyperspectral, and radar remotely sensed data coupled with statistical or geostatistical methods. For instance, Tóth et al. [15] compared classification trees and multiple linear regression modeling for mapping salt-affected soils in Hungary. The results demonstrated that the classification tree method performed better with an overall precision of 91–96% for the salt-affected soils and 99% for the nonsalt-affected soils using soil texture and composition, groundwater physio-chemical properties. Bakacsi et al. [16] successfully assessed soil salinization risks under

* Corresponding author: Ghada Sahbeni, Department of Geophysics and Space Science, Eötvös Loránd University, Pázmány Péter stny. 1/A, Budapest, 1117, Hungary, e-mail: gsahbeni@caesar.elte.hu

irrigation on the national scale using data retrieved from MARTHA 1.0 (Hungarian Detailed Soil Hydraulic Database) and USDA database. The study used a series of auxiliary covariates, e.g., salinity and depth of groundwater, soil texture, and aridity index, proven to be correlated with salinity fluctuation levels. Decision-makers can use the final salinization vulnerability map to enhance land use management and sustainable agronomy. Szatmári et al. [17] used a combination of random forest and multivariate geostatistical techniques and spectral indices derived from Sentinel-2 MSI images to predict the spatial distribution of salt-affected soil indicators, contributing to an international effort to create a global map of salt-affected soils (GSSmap). The findings demonstrated the significance of satellite-derived spectral indices and climatic, geomorphometric, and legacy data in modeling soil salinity with acceptable accuracy.

In Algeria, Dehni and Lounis [18] revealed the importance of multitemporal images of Landsat in identifying and delineating saline and sodic soils. Zurqani et al. [19] used a time series of remotely sensed data for 29 years and site observations to detect the spatiotemporal variation of soil salinity in Libya. In Uzbekistan, Ivushkin et al. [20] found a significant correlation between soil salinity and canopy temperature index retrieved from a moderate resolution imaging spectroradiometer (MODIS) image. Using SAR Sentinel-1 Data and Advanced Machine Learning Algorithms, the Gaussian processes method yielded a high performance in explaining the relationship between remotely sensed data and electrical conductivity (EC) field measurements in Vietnam, with a coefficient of determination equals 0.808 [21]. A study by Morgan et al. [22] had combined shortwave infrared band, the normalized difference vegetation index (NDVI), and the second PCA derived from Sentinel 2 MSI data with artificial neural networks to produce a highly accurate model with an overall accuracy of 94% between the actual and predicted soil salinity values. Fan et al. [23] proved the effectiveness of partial least square regression (PLSR) modeling in salinity prediction when applied between soil salinity and ALI-convolved field spectra. The final model explains 74.9% of soil salinity spatial variance. El-Battay et al. [24] created Soil Salinity and Sodicity Index 2 (SSSI₂) based on a semi-empirical model ability to differentiate between high salinity and extreme salinity areas with an accuracy ranges between 30–75% dS/m. Mousavi et al. [25] revealed the outperformance of the artificial neural network in soil salinity prediction ($R^2 = 0.964$ and $RMSE = 2.237$) when compared with multiple linear regression modeling ($R^2 = 0.506$ and $RMSE = 9.674$). Tajgardan et al. [26] assessed the performance

of geostatistical and statistical models, i.e., ordinary Kriging (OK), regression Kriging (RK), co-Kriging (CK), and multiple linear regression (MLR) to predict soil salinity in an arid area in northern Iran. The study revealed that the regression Kriging approach yielded the best performance among the applied approaches.

Despite the diversity of the studies mentioned earlier, they highlighted the effectiveness of remotely sensed data coupled with field measurements in mapping the spatio-temporal distribution of soil salinity. However, previous attempts to use regression analysis to model the relationship between soil salinity and remotely sensed data have been reasonably successful in many studies, with moderate accuracy and a high prediction error. Accordingly, geostatistical and statistical approaches open the doors to exploring soil parameters prediction possibilities with low cost and considerable accuracy.

The aims of this study are as follows: (i) to study the potential of Sentinel-2 MSI data in soil salinity prediction, (ii) to highlight the importance of spectral enhancements in salts detection, and (iii) to model the relationship between field measurements and remotely sensed data using a partial least squares regression (PLSR) approach.

2 Study area

The study area is located in the Great Hungarian Plain (GHP), lying approximately between latitudes 46°52' and 47°49' and longitudes 21°2' and 22°4' (Figure 1). It covers an area of 6903.5 km², with an average elevation of 88.8 m above sea level. The Tisza is the main river that runs through the plain, collecting tributaries from nearly the lowland region. Three kinds of deposits exist on the landscape: wind-blown sand on higher elevation areas, loess and loess-like sediments above the floodplain level, and silty clay in flat alluvial areas [27]. A moderately warm-dry climate characterizes the region with a yearly precipitation of around 560 mm and average annual evaporation of 900 mm. The mean annual temperature equals 11°C [28,29]. The rainiest month is June, and the least rainy is January, with an average precipitation of 71 and 30 mm, respectively [15].

3 Soil sampling and data collection

Eighty-one soil samples were selected within the upper 30 cm soil layer in the study area, with 74 of them as non-

Figure 1: Location of the study area: Sentinel-2 MSI image (True Color Composite) acquired on August 28, 2016.

saline ($SSC < 1 \text{ g/kg}$), four as low saline ($1 \text{ g/kg} < SSC < 2 \text{ g/kg}$), two as moderately saline ($2 \text{ g/kg} < SSC < 4 \text{ g/kg}$), and one as highly saline ($SSC > 4 \text{ g/kg}$). The soil sampling campaign is conducted between September 15 and October 15 to better detect salts' spectral characteristics during their accumulation [30,31] in the Hungarian Soil Information and Monitoring System (SIMS) framework. It is a nationwide soil monitoring program that provides soil information from around 1,235 locations, and it is considered the most unified and thematically detailed soil database in Hungary [30]. Regional soil experts selected sampling sites based on available soil information and their local experience. An average sample is collected from nine drillings from the 0–30, 30–60, and 60–90 cm soil layers in a 50 m diameter circle [31,32]. In the laboratory, saturated paste's resistance and conductivity are measured in the Hungarian Standard MSZ-08-0206/2-1978 [32,33]. The electrical resistance is measured by immersing an electrode in the water-saturated soil paste at the upper limit of the plasticity [34]. Details about MSZ-08-0206/2-1978 can be found in (MSZ 1978, 1978) [35].

Sentinel 2 MSI Level-1C (TOA) image was acquired from the European Space Agency (ESA) Copernicus portal.

Due to the scarcity of cloudless data (cloud cover 5%), satellite image acquisition date has a few days gap with the field measurements date. Yet, we assumed that soil salinity is relatively more stable than weather variations in that period of the year. Nonetheless, we geo-rectified the data to the Universal Transverse Mercator (UTM) coordinate system using World Geodetic System (WGS) 1984 datum assigned to north UTM Zone 34. These data were downloaded in the Geographic Markup Language JPEG2000 (GMLJP2) format. Level-1C TOA reflectance was radiometrically corrected using a Sen2Cor-SNAP processor to a level-2A TOA reflectance product [36].

The MultiSpectral Instrument (MSI) sensor includes 13 spectral bands from visible to shortwave infrared with a spatial resolution ranges from 10 to 60 m and a radiometric resolution of 12 bit. The list of Sentinel 2 MSI bands with their bandwidth, central wavelengths, and resolution is presented in Table 1. Many authors have proved that using remote sensing in the visible, near-infrared, and shortwave-infrared spectral ranges produces valuable spatial information that can detect soil salinity [37–42]. On that account, the visible, near-infrared, and short-wavelength bands were integrated into the regression analysis (eight bands).

Table 1: Spectral bands range and spatial resolution of Sentinel-2A MSI (ESA, 2020)

Sentinel-2 bands	Bandwidth (nm)	Central wavelength (nm)	Resolution (m)
Band 1: Coastal aerosol	21	443	60
Band 2: Blue	66	490	10
Band 3: Green	36	560	10
Band 4: Red	31	665	10
Band 5: Vegetation red edge	15	705	20
Band 6: Vegetation red edge	15	740	20
Band 7: Vegetation red edge	20	783	20
Band 8: NIR	106	842	10
Band 8A: Vegetation red edge	21	865	20
Band 9: Water vapor	20	945	60
Band 10: SWIR – Cirrus	31	1,375	60
Band 11: SWIR	91	1,610	20
Band 12: SWIR	175	2,190	20

Source: <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>.

The topography effect in salt movement within the profile was discussed by many authors, including Refs. [43,44], in similar landscapes, where the negative water balance shown in closed plains encourages the accumulation of soluble materials. Therefore, we obtained a 30 m SRTM Digital Elevation Model (DEM) in the GeoTIFF format from the OpenTopography Facility with support from the National Science Foundation. The auxiliary terrain variable was included in the analysis to investigate its role in salinity spatial distribution.

4 Methodology

Many scholars [45–51] have suggested the efficiency of spectral indices mentioned in Table 2 in soil salinity detection. Therefore, the reflectance of corresponding bands was used to calculate the spectral indices that relate vegetation performance to soil salinity, including NDVI, SAVI, SIs, and brightness index (BI).

Figure 2 illustrates the methodology of this study.

Linear regression is a simple algorithm that can be trained quickly and efficiently on systems with limited computational resources. Meanwhile, its computational complexity is lower than other machine learning algorithms [59,60]. Equation (1) shows the relationship between dependent and explanatory variables in its linear form.

$$Y = A_0 + A_1 * x_1 + A_2 * x_2 + A_3 * x_3 + \dots A_n * x_n, \quad (1)$$

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

We conducted a multivariate regression analysis to reveal the potential association between remotely sensed data and field measurements. Then, the Akaike information criterion (AIC) method was adopted to select the most significant variables for our model. This method assesses a model's ability to balance the data set without overfitting [61]. The AIC score rewards models with a high

Table 2: Spectral indices and their mathematical expressions

Index	Expression
Normalized difference vegetation index (NDVI)	$(NIR - R)/(NIR + R)$ [52]
Normalized difference salinity index (NDSI)	$(R - NIR)/(R + NIR)$ [53]
Vegetation soil salinity index (VSSI)	$2 * G - 5 * (R + NIR)$ [18]
Brightness index (BI)	$\sqrt{(R^2 + NIR^2)}$ [53]
Salinity index (SI)	$(R * G)/B$ [54]
Salinity index 1 (SI ₁)	$\sqrt{(G * R)}$ [50]
Salinity index 2 (SI ₂)	$\sqrt{(R * NIR)}$ [18]
Salinity index 3 (SI ₃)	$\sqrt{(G^2 + R^2 + NIR^2)}$ [50]
Salinity index 4 (SI ₄)	$\sqrt{(G^2 + R^2)}$ [55]
Ratio vegetation index (RVI)	R/NIR [56]
Intensity index 1 (Int ₁)	$(G + R)/2$ [57]
Intensity index 2 (Int ₂)	$(G + R + NIR)/2$ [57]
Simple ratio (SR)	$(R - NIR)/(G + NIR)$ [18]
Soil adjusted vegetation index (SAVI)	$(1 + L) * (NIR - R)/(NIR + R + L)$ [58]
Soil salinity and sodicity index 1 (SSSI ₁)	$SWIR_1 - SWIR_2$ [45]
Soil salinity and sodicity index 2 (SSSI ₂)	$(SWIR_1 * SWIR_2 - SWIR_2 * SWIR_2)/SWIR_1$ [45]

Figure 2: Schematic diagram of the methodology.

goodness-of-fit score and penalizes those that become too complex. A model with a lower AIC score should be able to strike a better equilibrium between its capacity to fit the data set and its ability to avoid overfitting it [62]. The basic formula is presented in equation (2).

$$AIC = -2(\log\text{-likelihood}) + 2K, \quad (2)$$

where log-likelihood is a measure of model fit and K is the number of model parameters.

Variables multicollinearity refers to the lack of independence of predictor variables in a regression-type analysis. It is considered problematic for parameter estimation as it increases the variance of regression parameters, leading to incorrect detection of significant predictors [63]. The partial least squares regression (PLSR) robustness to high collinearity is well known and widely used [64]. It condenses the predictors into a smaller set of uncorrelated components before applying least squares regression to these components [65]. However, PLSR is more prone to uncertainty estimation since the predictors are often calculated with error [23,65]. In this study, salt content data were used to build a PLSR model for mapping spatial salinity. Substantially, we developed the retrieval model using spectral indices, original bands, and elevation selected through a multiple linear regression followed by the AIC model selection approach. Statistically significant remotely sensed data were combined to estimate salt content. Details about PLSR method can be found in Haenlein and Kaplan [66] and Maitra and Yan [67]. Equation (3) shows its basic concept.

$$y = \sum_{i=1}^n w_i^* \rho_i + w_0, \quad (3)$$

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

Overall, the highest R^2 and the smallest RMSE indicate the best fit model. The RMSE is derived from equation (4).

$$RMSE = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2 / n}, \quad (4)$$

where n is the number of sampling sites, $\hat{y}_i$ is the estimated value at the point x_i , and y_i is the observed value at x_i .

R^2 measures the degree of statistical similarity between observed and expected values. The correlation is strong when $R^2 > 0.7$, moderate when $0.4 < R^2 < 0.7$, small when $0.2 < R^2 < 0.4$, and null when $R^2 < 0.2$. It is computed using equation (5).

$$R^2 = 1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (y_i - \bar{y})^2}, \quad (5)$$

where y is the actual value with a mean of $\bar{y}$, and $\hat{y}$ is the predicted value. The 5% probability level was adopted to test the correlation significance.

5 Results and discussion

The remarkable difference between a minimum of 0 g/kg of soil and a maximum of 5.6 g/kg indicates a spatial variability of salinity levels. Yet, the normality test demonstrated that salt content data had a positive-skewed distribution. This test result is confirmed by the disparity between a mean of 0.49 g/kg and a median of 0.3 g/kg. Thus, we used a square root transformation to normalize the data distribution, which was proved effective by skewness reduction from 4.14 for the initial data to 0.45 for the normalized data (Table 3).

Using R studio 4.0.2, a regression analysis was performed among the normalized salt content values and Sentinel 2 MSI bands, spectral indices, and elevation to explore the importance of these components. Then, we performed a PLSR to minimize the effect of multicollinearity between variables. The final model showed a great statistical significance with a p value of 0.000017 (< 0.05). Consequently, our approach yielded a highly

Table 3: Summary statistics of salt content samples (g/kg)

	Minimum	1st quantile	Median	Mean	3rd quantile	Maximum	Skewness
SSC	0	0.2	0.3	0.49	0.6	5.6	4.14

moderate correlation with a coefficient of determination R^2 equal to 0.68 and an RMSE equal to 0.22.

Figure 3 illustrates the relationship between measured and estimated normalized salt content values for the training set (60%) and the test set (40%). In many cases, the estimated salinity is higher than the field measurements. This issue was exposed in Figure 3 by an overestimation or underestimation of predictions. The abundance of null values, combined with the database's small size, resulted in slight inflation.

We retrieved the regression equation from R studio to create a soil salinity prediction map using the ArcMap 10.3 raster calculator. The linear relationship between field measurements and remotely sensed data is given by equation (6).

$$\begin{aligned}
 \sqrt{\text{SSC (g/kg)}} = & 0.4 - 1.31 \times \text{VSSI} - 50.17 \times \text{SI}_3 \\
 & - 52.42 \times \text{int}_1 + 87.01 \times \text{int}_2 \\
 & + 8.75 \times \text{SSSI}_1 - 22.4 \times \text{SSSI}_2 \\
 & + 0.0016 \times B_1 + 0.0068 \times B_2 \\
 & - 0.0044 \times B_3 - 0.0007 \times B_{11} \\
 & - 0.0006 \times B_{12} - 0.89 \\
 & \times \log(\text{elevation}),
 \end{aligned} \quad (6)$$

Figure 3: Relationship between measured and predicted normalized salt content values.

In order to identify soil salinity levels, we classified the pixels into five classes based on the Chinese classification scheme (Chinese Academy of Sciences, 2001), where the weight of salt content is measured per unit kg of soil as presented in Table 4 [68].

Figure 4 shows that approximately 72% of the pixels are classified as non-saline soils, 25% as low saline soils, and roughly 3% as moderately saline soils. The prediction map can outline soil salinity levels, provide further assistance in soil management strategies for decision makers, and elucidate the possibilities of using remote sensing and GIS tools in soil monitoring.

The model showed a highly moderate accuracy ($R^2 = 0.68$) at a probability level of 95% and an RMSE of 0.22. This approach has a remarkable efficiency in predicting and mapping the spatial changes of soil salinity. This is mainly due to the salinity indicators selected by the regression analysis and then the application of PLSR. In fact, the model revealed a superiority in terms of predictive performance comparing to the models reported by Mousavi et al. [25] ($R^2 = 0.506$), Shrestha [69] ($R^2 = 0.23$), Shamsi et al. [70] ($R^2 = 0.39$), and Hihi et al. [71] ($R^2 = 0.48$). Furthermore, spectral transformations, i.e., spectral indices, principal component analysis, can yield acceptable results in terms of soil salinity retrieval. The study by Allbed et al. [54] has shown that using spectral indices, i.e., salinity index (SI) and IKONOS original bands combined with field measurements of electrical conductivity (EC), produced a moderate correlation with a coefficient of determination R^2 equals 0.65. Yildirim et al. [72] demonstrated that soil salinity indices derived from Sentinel 2 MSI visible spectral bands produced better results than infrared spectra-based salinity indices for salinity detection.

Moreover, soil physical and chemical properties, mineralogy of salt crystals, color, and surface roughness influence the soil spectral reflectance in the visible and near-infrared spectra where salty surfaces appear bright crusts thick or with puffy structures. This adopted approach demonstrates the superiority of spectral indices derived from visible red and NIR bands over other covariates in soil salinity modeling. This fact was supported by the studies of Allbed et al. [54], Yildirim et al. [72], and Mehta et al. [73].

Table 4: Salinity content levels using the saline soil classification standard of China

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.

Figure 4: Soil salinity prediction map using the final model.

In addition, this study explores the shortwave infrared spectral bands’ significance in soil salinity detection, agreeing with the studies by Hihi et al. [71] and Lamqadem et al. [74]. Bannari et al. [75] proved the superiority of Sentinel-MSI SWIR bands in soil salinity modeling throughout laboratory electrical conductivity (EC_{Lab}) that showed a moderate correlation with SWIR bands (R^2 of 50% for $SWIR_1$ and 64% for $SWIR_2$).

The blue channel has also revealed a remarkable correlation with salts detection, which agrees with the findings of Mousavi et al. [25]. In the same context, Metternicht and Zink [40] proved that salt has a high spectral reflectance in the visible spectra, particularly in the blue band.

Overall, combining multispectral response derivatives produces more significant results for soil salinity modeling than only original bands, which is supported

by the studies of Eldeiry and Garcia [76], Noroozi et al. [77], Zewdu et al. [78], and Sahbeni [79].

Partial least squares regression enables the merging of spectral bands with spectral enhancements originating from the same image into a single model, reducing the issue of collinearity that exists in traditional regression analysis. These results are supported by Yu et al. [80], who used Landsat 8 OLI imagery and PLSR modeling to map soil salinity in China’s west Jilin Province. The findings showed that SI_3 and the Green channel were the most influential variables, with a final model characterized by a coefficient of determination R^2 equal to 0.698.

Our model successfully explained 68% of the data spatial variance, proving the potential usefulness of multispectral imagery coupled with regression analysis in soil salinity estimation. Although the primary goal of environmental modeling is to use existing relationships between covariates to simplify complex processes, the association between salinity and multispectral data has been entirely determined neither by this work nor by other previous studies, creating a space for open questions about the nature of this relationship and how to explain it. This topic will be discussed in future research.

6 Conclusion

This study demonstrates the possibility of mapping and quantifying spatial changes in soil salinity on a regional scale based on statistical modeling. By using Sentinel 2 MSI bands, spectral indices, and elevation, PLSR provides an acceptable and affordable approach to predict soil salinity with highly moderate accuracy. The model successfully explained 68% of the spatial data variance with a high statistical significance in a probability level of 5%. Therefore, the prediction map can outline salinity levels and assist decision makers in land use management. However, several limitations are given attention, i.e., the database’s small size affected the model accuracy as a minimal representation of each soil class/type is generally required to detect the complexity to be mapped. In addition, the moderate correlation between salt content and independent variables with a significant

prediction error raises interest regarding the essence of their relationship, especially that the soil is a highly variable continuum. Merely, these issues can be addressed when a more extensive database size is available and additional covariates are ready to be integrated into the model.

In future work, we will explore more robust geostatistical and statistical approaches that can yield higher accuracy and sketch the possible nonlinear nature of this process, taking into consideration environmental factors, e.g., precipitation, temperature, groundwater salinity levels, and other data types such as land use maps and legacy data.

Acknowledgments: The author would like to genuinely thank Prof. László Pásztor, from *Magyar Tudományos Akadémia Agrártudományi Kutatóközpont Talajtani és Agrokémiai Intézet*, who provided the soil sampling results, the anonymous reviewers and the editor for their constructive comments that helped to improve the manuscript quality.

Funding information: The APC for OA publication are covered by Eötvös Loránd University.

Conflict of interest: Author states no conflict of interest.

References

- [1] Asfaw E, Suryabagavan KV, Argaw M. Soil salinity modeling and mapping using remote sensing and GIS: the case of Wonji sugar cane irrigation farm, Ethiopia. *J Saudi Soc Agric Sci*. 2018;17(3):250–8. doi: 10.1016/j.jssas.2016.05.003.
- [2] FAO. Extend and causes of salt-affected soils in participating countries, Global network on integrated soil management for sustainable use of salt-affected soils; 2000. FAO-AGL website.
- [3] Tóth T. Monitoring, predicting and quantifying soil salinity, sodicity and alkalinity in Hungary at different scale. Past experiences, current achievements and an outlook with special regard to European Union; 1991.
- [4] Scudiero E, Corwin DL, Anderson RG, Yemoto K, Clary W, Wang Z, et al. Remote sensing is a viable tool for mapping soil salinity in agricultural lands. *Calif Agric*. 2017;71(4):231–8. doi: 10.3733/ca.2017a0009.
- [5] Pásztor L. Advanced GIS and RS applications for soil and land degradation assessment and mapping. *ISPRS Int J Geo-Inf*. 2021;10:128. doi: 10.3390/ijgi10030128.
- [6] Ben-Dor E, Goldshleger N, Benyamini Y, Agassi M, Blumberg DG. The spectral reflectance properties of soil structural crusts in the 1.2- to 2.5- μ m spectral region. *Soil Sci Soc Am J*. 2003;67(1):289. doi: 10.2136/sssaj2003.0289.
- [7] Aceves EÁ, Guevara HJP, Enríquez AC, Gaxiola JDC, Cervantes MDJP, Barrientos JH, et al. Determining salinity and ion soil using satellite image processing. *Pol J Environ Stud*. 2019;28(3):1549–60. doi: 10.15244/pjoes/81693.
- [8] Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. Proximal sensing of soil pH and lime requirement by mid infrared diffuse reflectance spectroscopy. In: Grenier G, Blackmore S, Steffe J, editors. 3 ECPA-EFITA Proceedings: Third European Conference on Precision Agriculture, Montpellier; 2001:1. p. 497–502. Agro Montpellier.
- [9] Metternicht GI. Analysing the relationship between ground-based reflectance and environmental indicators of salinity processes in the Cochabamba valleys (Bolivia). *Int J Ecol Environ Sci*. 1998;24(4):359–70.
- [10] Szatmári G, Pásztor L. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma*. 2019;337:1329–40. doi: 10.1016/j.geoderma.2018.09.008.
- [11] Csillag F, Pásztor L, Biehl LL. Spectral band selection for the characterization of salinity status of soils. *Remote Sens Environ*. 1993;43(3):231–42. doi: 10.1016/0034-4257(93)90068-9.
- [12] Sahu SK, Prasad MBNV, Tripathy BK. PCA Classification Technique of Remote Sensing Analysis of Colour Composite Image of Chillika Lagoon, Odisha. *Int J Adv Res Computer Sci Softw Eng*. 2015;5(5):513–8.
- [13] Li B, Ti C, Zhao Y, Yan X. Estimating soil moisture with Landsat data and its application in extracting the spatial distribution of winter flooded paddies. *Remote Sens*. 2016;8(1):38. doi: 10.3390/rs8010038.
- [14] Wu W. The generalized difference vegetation index (GDVI) for dryland characterization. *Remote Sens*. 2014;6(2):1211–33. doi: 10.3390/rs6021211.
- [15] Tóth T, Pásztor L, Kabos S, Kuti L. Statistical prediction of the presence of salt-affected soils by using digitalized hydrogeological maps. *Arid Land Res Manag*. 2002;16(1):55–68. doi: 10.1080/153249802753365322.
- [16] Bakacsi Z, Tóth T, Makó A, Barna G, Laborczi A, Szabó J, et al. National level assessment of soil salinization and structural degradation risks under irrigation. *Hung Geogr Bull*. 2019;68(2):141–56.
- [17] Szatmári G, Bakacsi Z, Laborczi A, Petrik O, Pataki R, Tóth T, et al. Elaborating Hungarian segment of the global map of salt-affected soils (GSSmap): national contribution to an international initiative. *Remote Sens*. 2020;12(24):4073. MDPI AG. doi: 10.3390/rs12244073.
- [18] Dehni A, Lounis M. Remote sensing techniques for salt-affected soil mapping: application to the Oran region of Algeria. *Procedia Eng*. 2012;33:188–98. doi: 10.1016/j.proeng.2012.01.1193.
- [19] Zurqani H, Nwer B, Rhoma E. Assessment of spatial and temporal variations of soil salinity using remote sensing and geographic information system in Libya. Singapore: Global Science and Technology Forum Pte Ltd; 2012. doi: 10.5176/2251-3361_geos12.64.
- [20] Ivushkin K, Bartholomeus H, Bregt AK, Pulatov A. Satellite thermography for soil salinity assessment of cropped areas in Uzbekistan Land Degradation and Development. 2017;28(3):870–7. doi: 10.1002/ldr.2670.
- [21] Hoa PV, Giang NV, Binh NA, Hai LVH, Pham TD, Hasanlou M, et al. Soil salinity mapping using SAR Sentinel-1 data and

- advanced machine learning algorithms: a case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sens.* 2019;11(2):128. doi: 10.3390/rs11020128.
- [22] Morgan RS, Abd El-Hady M, Rahim IS. Soil salinity mapping utilizing sentinel-2 and neural networks. *Indian J Agric Res.* 2018;52(5):524–9. ISSN:0367-8245. doi: 10.18805/IJARE.A-316.
- [23] Fan X, Yuanbo L, Jinmei T, Yongling W. Soil salinity retrieval from advanced multispectral sensor with partial least square regression. *Remote Sens.* 2015;7(1):488–511. doi: 10.3390/rs70100488.
- [24] El-Battay A, Bannari A, Hameid NA, Abahussain AA. Comparative study among different semi-empirical models for soil salinity prediction in an arid environment using OLI landsat-8 data. *Adv Remote Sens.* 2017;6(1):23–39. doi: 10.4236/ars.2017.61002.
- [25] Mousavi SZ, Habibnejad M, Kaviani A, Solaimani K, Khormali F. Digital mapping of topsoil salinity using remote sensing indices in Agh-Ghala Plain, Iran. *Ecopersia.* 2017;5(2):1771–86.
- [26] Tajgardan T, Ayoubi S, Shataee S, Sahrawat KL. Soil surface salinity prediction using aster data: comparing statistical and geostatistical models. *Aust J Basic Appl Sci.* 2010;4(3):457–67.
- [27] Ronai A. The quaternary of the great Hungarian plain. *Geologica Hungarica. Series geologica.* 1985;Tom. 21: 413–45. Retrieved from: http://epa.oszk.hu/02900/02986/00025/pdf/EPA02986_geologica_hungarica_ser_geol_1985_21_413-445.pdf.
- [28] Hungarian Meteorological Service. The climate of Hungary – general characteristics; 2021, Retrieved from: https://www.met.hu/en/eghajlat/magyarorszag_eghajlata/altalanos_eghajlati_jellemzes/altalanos_leiras/
- [29] Tóth T, Balog K, Szabo A, Pásztor L, Jobbágy EG, Nosetto MD, et al. 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.* 2014;6:1–15. doi: 10.1093/aobpla/plu054.
- [30] Szabó J, Pirkó B. The soil information and monitoring system (TIM); 2017. http://eagri.cz/public/web/file/519224/_2_7_Bela_Pirko.pdf.
- [31] Berényi-Üveges J, Szentes D, Talaj M, Rendszer M. A Talajmonitoring Rendszer Hungarian soil monitoring system soil monitoring system (TIM); 2016. https://esdac.jrc.ec.europa.eu/projects/esoter/Danube/Presentations/c%20-%20Hungary%20-%20Judit%20Berényi%20Üveges%20-%20TIM2015_Ispra.pdf.
- [32] Pásztor L. “Details about the protocol of soil sampling and salinity measurement”, email message to Ghada Sahbeni, Budapest, 2021 Jan 1.
- [33] Weynants M, Montanarella L, Toth G, Arnoldussen A, Anaya Romero M, Bilas G, et al. European hydrogeological data inventory (EU-HYDI). EUR 26053. Luxembourg (Luxembourg): Publications Office of the European Union; 2013. JRC81129.
- [34] Imre Kádár. 2 MANUAL remediation of contaminated soils assessment, Responsible publisher: Ministry of the Environment; 1998. Available from: <http://fava.hu/kvvm/www.kvvm.hu/szakmai/karmentes/kiadvanyok/karmkezikk2/2-09.htm>
- [35] MSZ 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).
- [36] Mueller-Wilm U. Sen2Cor Configuration, and User Manual V2.4. European Space Agency; 2017. p. 1–53. http://step.esa.int/thirdparties/sen2cor/2.4.0/Sen2Cor_240_Documentation_PDF/S2-PDGS-MPC-L2A-SUM-V2.4.0.pdf (accessed on May 20, 2021).
- [37] Mougenot B, Pouget M. Remote sensing of salt-affected soils. *Remote Sens Rev.* 1993;7:241–59.
- [38] Rahman H, Dedieu G. SMAC: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *Int J Remote Sens.* 1994;15:123–43.
- [39] Khan NM, Rastokuev VV, Shalina EV, Sato Y. Mapping salt-affected soils using remote sensing indicators – A simple approach with the use of GIS IDRISI. Proceedings of the 22nd Asian conference on Remote sensing, 5–9 November 2001. Singapore: Center for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore; Singapore Institute of Surveyors and Valuers; Asian Association on Remote Sensing.
- [40] Metternicht GI, Zinck JA. Spatial discrimination of salt- and sodium-affected soil surfaces. *Int J Remote Sens.* 1997;18(12):2571–86.
- [41] Yu R, Liu T, Xu Y, Zhu C, Zhang Q, Qu Z, et al. Analysis of salinization dynamics by remote sensing in Hetao Irrigation District of North China. *Agric Water Manag.* 2010;97:1952–60. doi: 10.1016/j.agwat.2010.03.009; Setia R, Lewis M, Marschner P, Raja Segaran R, Summers D, Chittleborough D. Severity of salinity accurately detected and classified on a paddock scale with high-resolution multispectral satellite imagery. *Land Degrad Dev.* 2010;24:375–84. doi: 10.1002/ldr.1134.
- [42] Sidike A, Zhao S, Wen Y. Estimating soil salinity in Pingluo County of China using QuickBird data and soil reflectance spectra. *Int J Appl Earth Obs Geoinf.* 2014;26:156–75. doi: 10.1016/j.jag.2013.06.002.
- [43] Kodva VV. Geochemistry of the Arid Zone in USSR. Moskva-Leningrad: Academy of Sciences; 1954.
- [44] Schofield R, Thomas DSG, Kirkby MJ. Causal processes of soil salinization in Tunisia, Spain, and Hungary. *Land Degrad Dev.* 2001;12(2):163–81.
- [45] Bannari A, Guedon AM, El-Harti A, Cherkaoui FZ, El-Gharni A. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO-1) sensor. *Commun Soil Sci Plant Anal.* 2008;39(19–20):2795–811.
- [46] Gorji T, Yildirim A, Sertel E, Tanik A. Remote sensing approaches and mapping methods for monitoring soil salinity under different climate regimes. *Int J Environ Geoinf.* 2019;6(1):33–49. doi: 10.30897/ijgeo.
- [47] Hernández EI, Melendez-Pastor I, Navarro-Pedreño J, Gómez I. Spectral indices for the detection of salinity effects in melon plants. *Sci Agricola.* 2014;71(4):324–30. doi: 10.1590/0103-9016-2013-0338.
- [48] Nouri H, Chavoshi Borujeni S, Alaghmand S, Anderson SJ, Sutton PC, Parvazian S, et al. Soil salinity mapping of urban greenery using remote sensing and proximal sensing techniques; the case of veale gardens within the Adelaide Parklands. *Sustainability.* 2018;10(8):2826. doi: 10.3390/su10082826.

- [49] Azabdaftari A, Sunar F. Soil salinity mapping using multitemporal Landsat data. *ISPRS Arch.* 2016;41:3–9. International Society for Photogrammetry and Remote Sensing. doi: 10.5194/isprsarchives-XLI-B7-3-2016.
- [50] Douaoui AEK, Nicolas H, Walter C. Detecting salinity hazards within a semi-arid context by means of combining soil and remote-sensing data. *Geoderma.* 2006;134(1–2):217–30. doi: 10.1016/j.geoderma.2005.10.009.
- [51] Guo B, Han B, Yang F, Fan Y, Jiang L, Chen S, et al. Salinization information extraction model based on VI–SI feature space combinations in the Yellow River Delta based on Landsat 8 OLI image. *Geomat Nat Hazards Risk.* 2019;10(1):1863–78. doi: 10.1080/19475705.2019.1650125.
- [52] Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden SC, Mercanti EP, Becker M, editors. *Third Earth Resources Technology Satellite–1 Symposium. Volume I: Technical Presentations.* NASA SP-351. Washington, DC: NASA; 1974. 309–17.
- [53] Khan NM, Rastokuev VV, Sato Y, Shiozawa S. Assessment of hydro saline land degradation by using a simple approach of remote sensing indicators. *Agric Water Manage.* 2005;77:96–109.
- [54] Allbed A, Kumar L, Sinha P. Mapping and modelling spatial variation in soil salinity in the Al Hassa Oasis based on remote sensing indicators and regression techniques. *Remote Sens.* 2014;6(2):1137–57. doi: 10.3390/rs6021137.
- [55] Yahiaoui I, Douaoui A, Zhang Q, Ziane A. Soil salinity prediction in the Lower Cheliff plain (Algeria) based on remote sensing and topographic feature analysis. *J Arid Land.* 2015;7:794–805. doi: 10.1007/s40333-015-0053-9.
- [56] Krtalic A, Prodan A, Racetin I. Analysis of burned vegetation recovery by means of vegetation indices. 19th international multidisciplinary scientific geoconference SGEM 2019; Conference Proceedings, Volume 19; Informatics, geoinformatics and remote sensing, Issue: 2.2, Geodesy and mine surveying, photogrammetry and remote sensing, cartography and GIS/ – Sofia. Bulgaria: STEF92 Technology; 2019. p. 449–456. doi: 10.5593/sgem2019/2.2/s10.055
- [57] Bouaziz M, Matschullat J, Gloaguen R. Improved remote sensing detection of soil salinity from a semi-arid climate in Northeast Brazil. *Comptes Rendus Geosci.* 2011;343:795–803.
- [58] Huete AR. A soil-adjusted vegetation index (SAVI). *Remote Sens Environ.* 1988;25(3):295–309. doi: 10.1016/0034-4257(88)90106-X.
- [59] Mansouri E, Feizi F, Jafari Rad A, Arian M. 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.* 2018;9:373–84. doi: 10.5194/se-9-373-2018.
- [60] Wicki A, Parlow E. Multiple regression analysis for unmixing of surface temperature data in an urban environment. *Remote Sens.* 2017;9(7):684. doi: 10.3390/rs9070684.
- [61] Hirotugu Akaike A. New look at the statistical model identification. *IEEE Trans Autom Control.* 1974;19:716–23.
- [62] Fabozzi FJ, Focardi SM, Rachev ST, Arshanapalli BG. Appendix E: model selection criterion: AIC and BIC. In: Fabozzi FJ, Focardi SM, Rachev ST, Arshanapalli BG, editors. *The Basics of Financial Econometrics*; 2014. doi: 10.1002/9781118856406.app5.
- [63] Dormann C, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography.* 2013;36:27–46. doi: 10.1111/j.1600-0587.2012.07348.x.
- [64] Cai T, Cunyong J, Yang X. 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.* 2009;23(3):248–61. doi: 10.1080/15324980903038701.
- [65] Cho MA, Skidmore A, Corsi F, van Wieren SE, Sobhan I. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *Int J Appl Earth Obs Geoinf.* 2007;9:414–24.
- [66] Haenlein M, Kaplan A. A beginner's guide to partial least squares analysis. *Underst Stat.* 2004;3:283–297. doi: 10.1207/s15328031us0304_4.
- [67] Maitra S, Yan J. Principle component analysis and partial least squares: two dimension reduction techniques for regression; 2008. https://www.casact.org/sites/default/files/database/dpp_dpp08_08dpp76.pdf.
- [68] Weng YL, Gong P, Zhu ZL. A spectral index for estimating soil salinity in the Yellow River delta region of China using EO-1 hyperion data. *Pedosphere.* 2010;20(3):378–88. doi: 10.1016/S1002-0160(10)60027-6.
- [69] Shrestha R. Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand. *Land Degrad Dev.* 2006;17:677–89.
- [70] Shamsi FRS, Sanaz Z, Abtahi AS. Soil salinity characteristics using moderate resolution imaging spectroradiometer (MODIS) images and statistical analysis. *Arch Agron Soil Sci.* 2013;59:471–89.
- [71] Hihi S, Rabah ZB, Bouaziz M, Chtourou MY, Bouaziz S. Prediction of soil salinity using remote sensing tools and linear regression model. *Adv Remote Sens.* 2019;08(03):77–88. doi: 10.4236/ars.2019.83005.
- [72] Yildirim A, Gorji T, Hamzehpour N, Sertel E. Comparison of different soil salinity indices derived from sentinel-2A images. *Int Symposium Appl Geoinformatics.* 2019;1(1):230/614. doi: ISBN: 978-975-461-564-7/2019.
- [73] Mehta M, Anh VL, Saha SK, Agrawal S. Evaluation of indices and parameters obtained from optical and thermal bands of landsat 7 ETM + for mapping of salt-affected soils and water-logged areas. *Asian J Geoinf.* 2012;12(4):9–16.
- [74] Lamqadem A, Saber H, Rahimi A. Mapping soil salinity using Sentinel-2 image in Ktaoua oasis (Southeast of Morocco), 7th Digital Earth Summit 2018. *El Jadida, Morocco: 2018 April 17–19.*
- [75] Bannari A, El-Battay A, Bannari R, Rhinane H. Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape. *Remote Sens.* 2018;10:855.
- [76] Eldeiry AA, Garcia LA. Detecting soil salinity in alfalfa fields using spatial modeling and remote sensing. *Soil Sci Soc Am J.* 2008;72:201–11. doi: 10.2136/sssaj2007.0013.
- [77] Noroozi AA, Homaei M, Farshad A. Integrated application of remote sensing and spatial statistical models to the

- identification of soil salinity: a case study from Garmsar Plain, Iran. *Env Sci.* 2012;9(1):59–74.
- [78] Zewdu S, Suryabhagavan KV, Balakrishnan M. Land-use/land-cover dynamics in Sego Irrigation Farm, southern Ethiopia: a comparison of temporal soil salinization using geospatial tools. *J Saudi Soc Agric Sci.* 2016;15(1):91–7. ISSN 1658-077X. doi: 10.1016/j.jssas.2014.03.003.
- [79] Sahbeni G. Soil salinity mapping using Landsat 8 OLI data and regression modeling in the Great Hungarian Plain. *SN Appl Sci.* 2021;3(587):1–13. doi: 10.1007/s42452-021-04587-4.
- [80] Yu H, Liu M, Du B, Wang Z, Hu L, Zhang B. Mapping soil salinity/sodicity by using landsat OLI imagery and PLSR algorithm over Semiarid West Jilin Province, China. *Sensors.* 2018;18:1048.