



Digital Mapping Algorithms to Estimate Soil Salinity in Indira Gandhi Nahar Pariyojana (IGNP) Command area of India

P. C. Moharana^{1*}, S. Dharumarajan², Nirmal Kumar¹, U. K. Pradhan³, R. K. Jena⁴,
R. K. Naitam¹, Sunil Kumar⁵, R. S. Singh⁶, R. S. Meena⁶, M. Nogiya⁶, R. L. Meena⁶
and B. L. Tailor⁶

¹ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur-440033

²ICAR-National Bureau of Soil Survey and Land Use Planning Regional Centre, Hebbal, Bangalore-560024

³ICAR-Indian Agricultural Statistics Research Institute, New Delhi -110 012

⁴ICAR-Indian Institute of Water Management, Bhubaneswar -751 023

⁵ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, New Delhi-110 012

⁶ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Udaipur-313001

Abstract: In the present study, the distribution of salinity was investigated using digital soil mapping (DSM) algorithms in the 5 km buffer zone of both sides of the Indira Gandhi Nahar Pariyojana (IGNP) canal system of Suratgarh tehsil in Rajasthan. To achieve this goal, 64 soil samples were used with 21 environmental covariates and 3 DSM algorithms. The result from the study showed that the difference between the minimum and maximum EC_e is very high ($35.55 dS m^{-1}$) in the different irrigation zone of the IGNP canal system. The EC_e ranged from 0.50 to 36.05 $dS m^{-1}$. Results indicated that the most important environmental covariates were annual precipitation, elevation and valley depth. Among the DSM algorithms, RF model showed the best performance in predicting EC_e at the regional level. Results showed that the RF algorithm could predict EC_e with an R^2 , RMSE and MAE of 0.701, 3.367 and 1.722, respectively. RF and QRF showed similar performance in predicting EC_e , while SVM showed lower efficiency than the other models in terms of R^2 and prediction errors. The salinity prediction map shows that the vulnerability to soil salinity is high in the Anupgarh branch of the canal, and low in the IGNP main and Bikaner canal area. Furthermore, the model developed in this study provides comprehensive guidance for the land planners and decision-makers to develop amicable strategies for the management of the IGNP canal system.

Keywords: Digital soil mapping, soil salinity, spatial distribution, IGNP command area, arid ecosystem

Introduction

Soil salinization is a worldwide problem, particularly in arid and semiarid regions with low rainfall and high evapotranspiration. Soil salinization is increasing at an alarming rate, and it is recognized as a global environmental issue with reports from all over the world (Gorji *et al.* 2017). The high evaporation conditions that exist in arid regions cause a steady rise in

salt content in water sources as well as in the soil (Singh 2015). In dry and semiarid regions, roughly 20% of irrigated land is salt-affected globally, and the land area with secondary soil salinization concerns might be as high as 80 Mha (million hectares) (Wang *et al.* 2018). Soil salinity and alkalinity impact 8.4 Mha in India alone, with roughly 5.5 Mha being waterlogged (Jaglan and Qureshi 1996). The Indira Gandhi Nahar Pariyojana (IGNP) is one of the most ambitious surface irrigation projects which

*Corresponding author: (Email: pravashiari@gmail.com)

occupy the north-western and far western parts of the Thar Desert of Rajasthan. The irrigation network has converted barren deserts into rich and green farmland since its inception in 1960. Salinity development and waterlogging, on the other hand, have become severe issues in the IGNP command area. Secondary salinization and salt build-up have occurred as a result of high temperatures, the existence of excessive water-soluble salts in soils, and a high rate of evapotranspiration. There have been several attempts to assess the extent and distribution of salt-affected soils in India. However, there is a paucity of data on the spatial prediction of soil salinity using a digital mapping methodology. As a result, in order to address these issues, an appropriate model for the spatial distribution of soil salinity must be evaluated.

Digital soil mapping (DSM) has now been widely used globally for mapping soil classes and properties (Arrouays *et al.* 2014). In particular, DSM has been used to map soil salinity efficiently around the world. DSM methodology has been adopted by FAO and ITPS (2020) so that digital soil maps can be produced reliably for sustainable land management. Mapping soil spatial variations by traditional field surveys are time-consuming and expensive, especially at national, regional or global scales (McBratney *et al.* 2003; Minasny *et al.* 2013). Therefore, it is necessary to have robust methods and models to predict soil properties at a given location or scale. Considerable advances in remote sensing techniques and machine learning approaches have allowed accurate prediction of soil properties with new methods like digital soil mapping (McBratney *et al.* 2003; Hengl *et al.* 2015). In recent years, DSM techniques have been used to map soil properties using environmental variables. These methods were designed to overcome the limitations of the conventional soil mapping approach and to estimate soil properties based on relationships between soil and environmental variables obtained from terrain attributes and satellite imagery (McBratney *et al.* 2003; Minasny and Hartemink 2011). In India, ICAR-National Bureau of Soil Survey and Land Use Planning (ICAR-NBSS&LUP), Nagpur has recently launched an

ambitious program called “Indian Soil Grids” with the objective to develop soil properties map as per Global Soil Map specifications (Dharumarajan *et al.* 2019).

As there is a lack of information about the spatial prediction of soil salinity in the IGNP command area of India, this study predicts the soil salinity spatially through the random forest (RF), quantile regression forest (QRF) and support vector machine (SVM) model techniques. The objectives of this study were to i) predict soil salinity with different digital mapping models, (ii) evaluate the effectiveness of models, and (iii) identify the most important environmental variables controlling the spatial distribution of soil salinity in the IGNP command area of India. We expect that our outputs would improve and update the current soil salinity information system with new fine-resolution soil salinity maps that could be useful to end-users and stakeholders.

Materials and Methods

Study site

The study area lies between 28.903°N to 29.686° N latitude and 73.441° E to 74.218° E longitude located in Suratgarh tehsil of Rajasthan, India covering 1,28,140 ha area (Fig. 1). This area has been irrigated by a canal network since the late 1960 and about 70% of cultivable fields of the studied area are under irrigation (Jaglan and Qureshi 1996). The study area is irrigated by the IGNP main canal and its branches like the Bikaner lift canal and Anupgarh canal, respectively under the IGNP irrigation project. The area has very scanty and erratic rainfall with extremely hot summer and cold winter. The average rainfall of the area is 286 mm. The dominant soils are deep to very deep, either calcareous or non-calcareous in nature. Texture varied from loamy sandy to sandy loam with weak structure. The soil belongs to Typic Haplocambids (Shyampura *et al.* 2002). Cereal and legume-based cropping systems are practised for more than the last 60 years. Crops included in the rotations are mustard (*Brassica juncea*), moth bean (*Vigna aconitifolia*), cluster bean (*Cyamopsis tetragonoloba*), groundnut (*Arachis hypogaea*), chickpea (*Cicer arietinum*), wheat

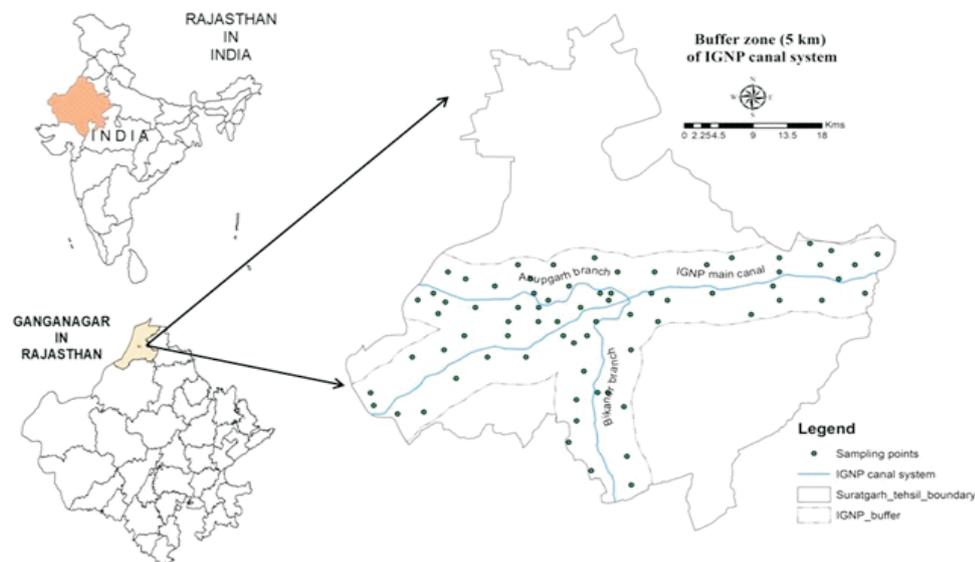


Fig. 1. Location map and soil sampling points of the IGNP canal system in the hot arid India

(*Triticum aestivum*), green gram (*Vigna radiata*), pearl millet (*Pennisetum glaucum*) and cotton (*Gossypium* sp.).

Soil sampling and analysis

Soil samples were collected in a 5 km buffer zone on both sides of the canal. Over 64 soil samples (arable layer) from different agricultural sites were collected during the period of March 2017. The geographical coordinates of the sampling points were recorded using a geographical positioning system (GPS). The samples were dried, ground and sieved with a 2.0 mm sieve and stored for analysis. The soil samples from different sites were analyzed for mechanical composition, pH, electric conductivity in the soil saturated paste extract (EC_e), calcium carbonate ($CaCO_3$), exchangeable cations (Exch. Ca, Mg, Na and K), exchangeable sodium percentage (ESP) following standard procedures. In this paper, EC_e data was used to estimate soil salinity prediction by digital mapping algorithms.

Environmental covariates

All terrain attributes used in this study were derived from a shuttle radar topographic mission (SRTM) DEM with a 30 m grid resolution. The primary and secondary derivatives of DEM like elevation, slope, aspect, curvatures (plan and profile), topographic wetness index (TWI), LS factor, Multi-resolution Ridge Top Flatness (MRRTF), Multi-resolution Index of Valley Bottom Flatness (MRVBF), Total Catchment Area (TCA) and valley depth were derived by using Saga-GIS 6.3.0 version (Table 1). Along with DEM attributes, other indices were also computed, which included salinity index (SI), soil adjusted vegetation index (SAVI), vegetation soil salinity index (VSSI), normalized difference salinity index (NDSI), normalized difference vegetation index (NDVI), salinity ratio (SR), and brightness index (BI) were also derived from the Landsat 8 satellite data. Three bioclimatic variables were selected for use as covariates in the DSM study and these variables are mean maximum temperature (Max Temp), mean minimum temperature (Min Temp) and annual precipitation (annual ppt). Raster data on bioclimatic

Table 1. Environmental covariates used for spatial prediction of soil EC_e

Environmental covariates	Surface parameters	Definition	Source	Soil forming factors	
Terrain attributes	Elevation	Height above sea level (m)	SRTM DEM	R	
	Aspect	The compass direction of the maximum rate of change	SRTM DEM	R	
	Slope	Average gradient above flow path	SRTM DEM	R	
	Multi-resolution Ridge-top Flatness Index (MRRTF)	Identifies the depositional areas	SRTM DEM	R	
	Multi-resolution Valley Bottom Flatness Index (MRVBF)	Identifies the depositional areas	SRTM DEM	R	
	LS factor	Slope-length factor	SRTM DEM	R	
	Plan curvature	Plan curvature is perpendicular to the slope	SRTM DEM	R	
	Profile curvature	Profile curvature is parallel to the direction of the maximum slope	SRTM DEM	R	
	Total catchment area (TCA)	Calculated the flow accumulation	SRTM DEM	R	
	Topographic wetness index (TWI)	TWI = $\ln (As/\tan \beta)$: where As is upslope catchment area and β is the slope gradient	SRTM DEM	R, CI	
	Valley depth	Valley depth	relative position of the valley (Metres)	SRTM DEM	R
		Salinity index (SI)	$SI = \sqrt{B4 * B3}$	Landsat 8	PM,S
	Remote sensing data	Salinity ratio (SR)	$SR = \frac{B3 - B4}{B2 + B4}$	Landsat 8	PM,S
Normalized difference salinity index (NDSI)		$NDSI = \frac{B4 - B5}{B4 + B5}$	Landsat 8	PM,S	
Normalized difference vegetation index (NDVI)		$NDVI = \frac{B5 - B4}{B5 + B4}$	Landsat 8	O	
Soil adjusted vegetation index (SAVI)		$SAVI = \frac{(B5 - B4)}{(B5 + B4 + 0.5)} * 1.5$	Landsat 8	PM,S	
Vegetation soil salinity index (VSSI)		$VSSI = 2 * B3 - 5(B4 + B5)$	Landsat 8	PM,S	
Climate		Brightness index (BI)	$BI = \sqrt{B3^2 + B4^2}$	Landsat 8	PM,S
		Min_Temp	Average min temperature	Worldclim2	CI
		Max_Temp	Average max temperature	Worldclim2	CI
		Annual_ppt	Precipitation	Worldclim2	CI

variables at 30s resolution were downloaded from <http://worldclim.org/current> for the whole world and the respective grids for the study area were extracted from these world grids.

Spatial prediction models

Random forest

RFs are a group of algorithms that have been developed as an extension of Classification and Regression Tree Analysis (CART) to enhance the prediction performance of the model (Breiman 2001). Number of tree (n tree), minimum no of samples at terminal node (n min) and a number of predictors used for fitting the tree (Mtry) are the three parameters that decide the fitting of the random forest model. The internal out-of-bag (OOB) prediction generated through boot strapping provides an estimate of accuracy across the decision trees which was used for the initial assessment of the performance of the model (Breiman 2001). RF is non-sensitive to missing data and has the capacity to handle a large number of both quantitative and categorical data (Dharumarajan *et al.* 2017). RFs have been mainly used for classification problems. RFs are a data-driven statistical approach that has recently been employed in DSM studies (Hengl *et al.* 2015). Given that it shows good accuracy, is fast, simple to use, has a useful internal estimation of error, and correlation and calculation of variable importance. For running the RF algorithms, the Random Forest package was used in the R environment.

Quantile regression forest

QRF model is an extension of the random forest model and the advantage of QRF over RF is for each node in each tree, RF keeps only the mean of the observations that fall into this node and neglects all other information whereas QRF keeps the value of all observations in this node, and assesses the conditional distribution based on the information (Meinshausen 2006; Dharumarajan *et al.* 2019). For the present study, the quantreg Forest package was used for running the QRF algorithm in an R environment.

Support vector machine

The SVM is a class of effective, very flexible modelling algorithms. The theory behind it was initially developed in the framework of classification models (Kuhn and Johnson 2013). SVM analysis proposed by Cortes and Vapnik (1995) is one of the general supervised machine learning tools for classification and regression. There are many planes in SVM to divide input data into different classes. Previous studies showed a better learning ability and lower prediction errors compared with many other algorithms (Maynard and Levi 2017; Mahmoudzadeh *et al.* 2020). For running the SVM algorithm, `e1071` and `caret` package were used.

Validation of statistical algorithms

The accuracy of the soil EC_e map was assessed through a cross-validation approach. For calibration of the model, 75% of soil datasets were used and the remaining 25% datasets were used for validation. To compare model performance, we use R^2 , RMSE and MAE (Dharumarajan *et al.* 2017;2019) described these measures of model performance as follows:

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{[\sum_{i=1}^n (O_i - \bar{O})^2 * \sum_{i=1}^n (P_i - \bar{P})^2]}} \right)^2$$

$$\text{Mean absolute error (MAE)} = \frac{1}{n} \sum_{i=1}^n |P_i - O_i|$$

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

where P_i and O_i are the predicted and observed EC_e ; n is the number of samples; \bar{P} and \bar{O} are the means for the predicted and observed EC_e . Models with the highest R^2 and lowest RMSE and MAE were deemed the best models. The model calibration and validation were repeated using different environmental variables.

Results and Discussion

Descriptive statistics

The descriptive statistics for EC_e are given in table 2. The results show that the mean EC_e in the study area is 4.60 dS m^{-1} . According to this table, the EC_e ranged from 0.50 to 36.05 dS m^{-1} in the different irrigation zone of the IGNP canal system. The distribution of EC_e in the Bikaner branch area ranged from 1.6 to 3.7 dS m^{-1} , which is not extra ordinarily high. In contrast, the EC_e values of the Anupgarh branch were higher than in IGNP main and Bikaner branch canal area (Fig. 2). The difference between the minimum and maximum EC_e is very high (35.55 dS m^{-1}), this difference showed that the surface soils are more easily affected by management, environmental variables and disturbance. The coefficient of variation (CV) for EC_e was high (more than 59.13%), which shows a wide range of values across the study area; it implies that the EC_e has strong spatial dependence. Wilding (1985) classified CV values into 3 classes with high ($CV > 35\%$), moderate (15%

$>CV < 35\%$) and low variability ($CV < 15\%$). According to CV analysis, our results indicate that EC_e has a semi-homogeneous spatial distribution that could be associated with the difference in land use, types of farming and management practices and topographic position in the study area.

Table 2. Summary statistics of soil salinity (EC_e) in the IGNP canal system.

Statistical parameter	$EC_e(\text{dS m}^{-1})$
Mean	4.60
Std. error of mean	0.90
Median	1.55
Std. deviation	7.17
CV	59.13
Skewness	2.72
Std. error of skewness	0.30
Kurtosis	7.38
Std. error of kurtosis	0.59
Minimum	0.50
Maximum	36.05
Percentiles	
	25 1.05
	50 1.55
	75 4.50

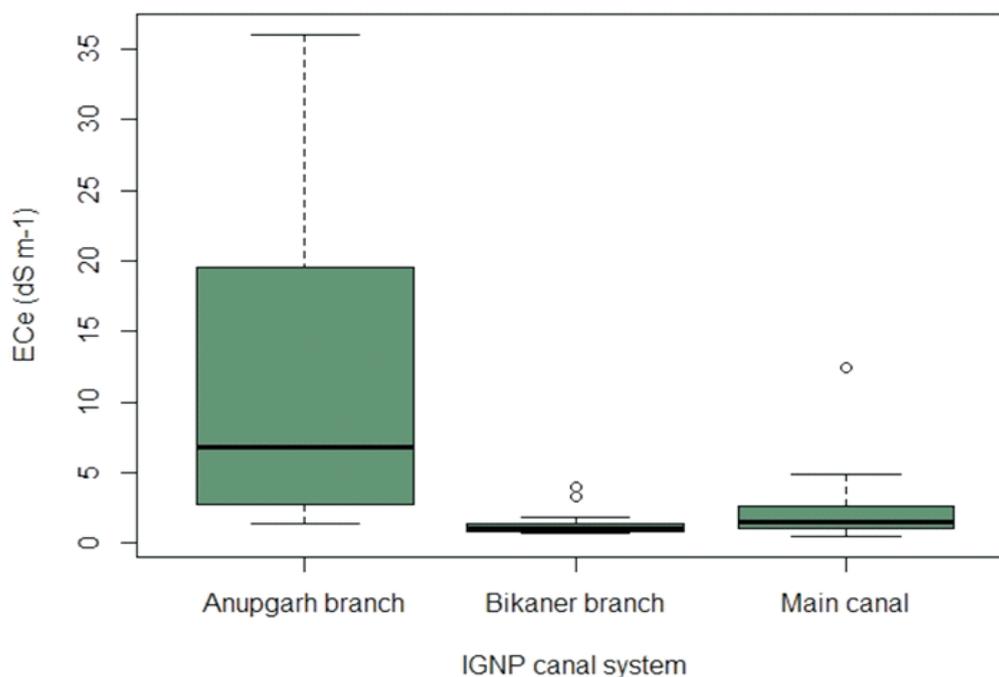
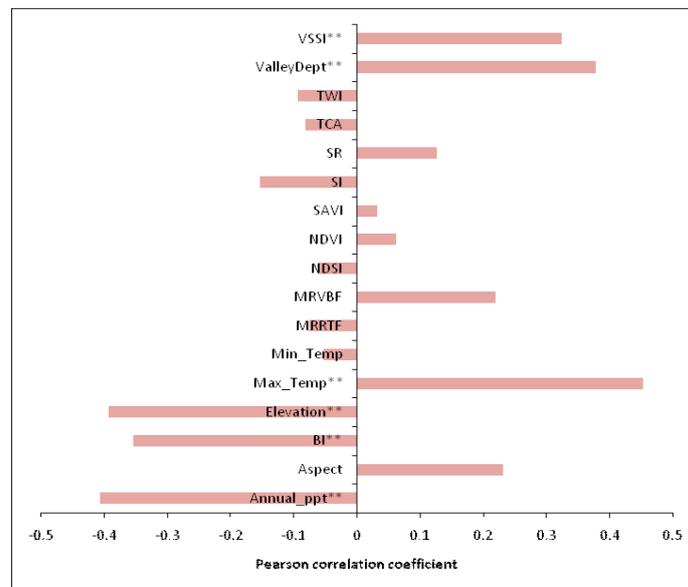


Fig. 2. Boxplot of EC_e under three irrigation zone of IGNP canal system.

Pearson correlation coefficients between EC_e and predictor variables are shown in fig.3. EC_e at the soil surface is positively correlated with max temp ($r = 0.452$) and negatively correlated with annual precipitation ($r = -0.406$) respectively. The correlation between climatic and topographic variables and the quantity of EC_e has been documented previously (Manickam *et al.* 2021). Climatic variables influence

soil properties, vegetation cover, water retention, soil erosion and also soil salinity. Our results also indicated a negative correlation between EC_e and the elevation ($r = -0.392$) and a positive correlation with valley depth ($r = 0.377$), similar to the findings of Taghizadeh-Mehrjardi *et al.* (2014). Other terrain attributes did not show a correlation with EC_e in our study area.



**Correlation is significant at the 0.01 level

Fig. 3. Pearson correlation between environmental covariates and EC_e

Relative importance of covariates

The parameter sensitivity analysis for the random forest model is presented in fig. 4. The variable importance is measured based on “out-of-bag” samples which mean observation is not included in the model. Another detail is that they are based on an MSE accuracy measure; in this case, the difference, when a covariate is included and excluded in a tree model. For the node purity, it is the total decrease in node impurities from splitting on the variable, averaged over all trees. Annual precipitation, elevation, aspect and valley depth are the best predictors to explain the variability of EC_e in our study area. In addition, the result reveals that the climatic

indicators contribute much more compared to terrain attributes in estimating salinity in an arid ecosystem. Kühn *et al.* (2009) reported that terrain attributes had weak and non-significant effects on the prediction of EC whereas geological unit and groundwater table map have a significant effect on the regression models.

Performance of DSM algorithms to predict EC_e

Comparing the statistical distribution and correlation between the models, we found that the higher Pearson correlation coefficient (r-value) between predicted values was between RF and QRF (0.99) (Fig. 5). We also found that the statistical distribution of predicted values is quite similar between the three

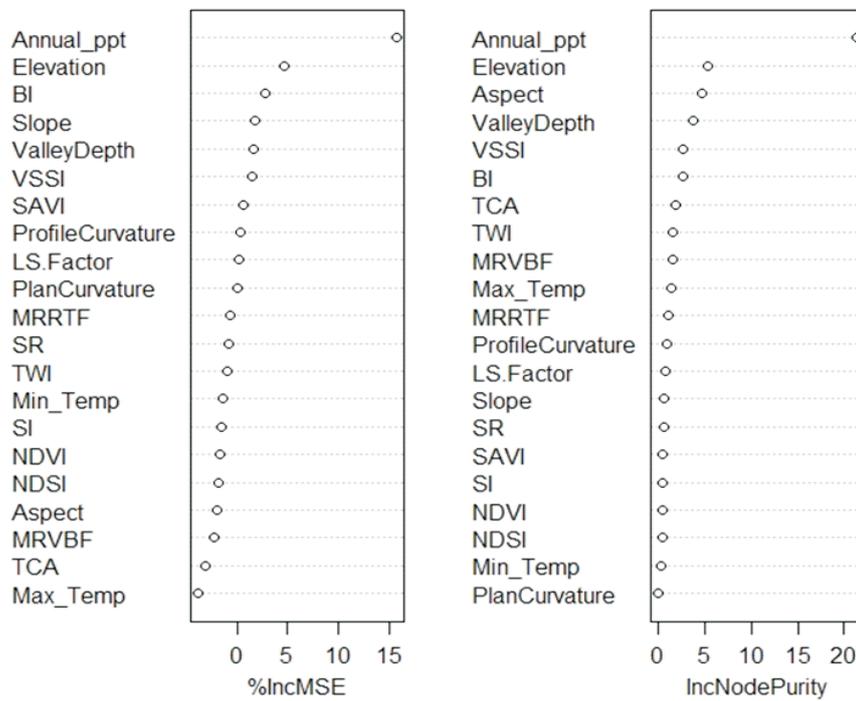


Fig. 4. Relative importance of environmental covariates using the best prediction model (RF).

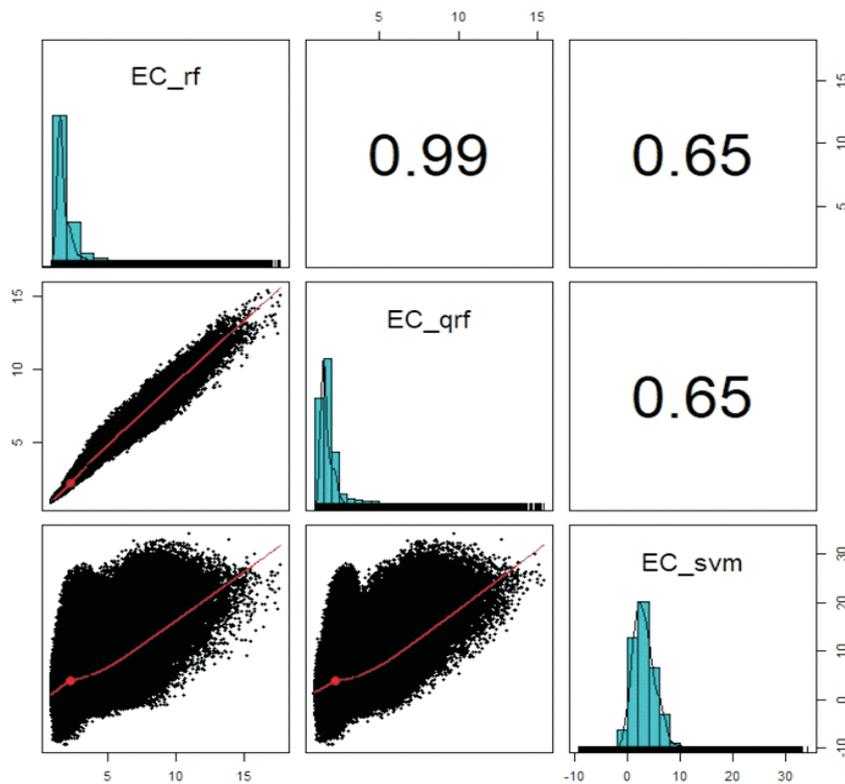


Fig. 5. Comparison of DSM model correlations (RF, QRF, SVM) and statistical distributions

methods and that the higher discrepancies were found in SVM model. The results of the cross-validation procedure used to compare the performance of RF, QRF and SVM are summarized in table 3 and fig. 6. Based on the error indices (R^2 , RMSE and MAE), all models performed well, but RF was best. We found that there were slight differences between RF and QRF models, suggesting that these two models were relatively stable in their predictive ability. The results showed that all models predicted EC_e acceptably well based on the range of average values of R^2 from 0.505 to 0.701; RMSE between 3.367 and 3.628 and MAE from 1.722 to 2.955. Results showed that the RF algorithm could predict EC_e with an R^2 , RMSE and MAE of 0.701, 3.367 and 1.722,

respectively. RF and QRF showed similar performance in predicting EC_e , while SVM showed lower efficiency than the other models in terms of R^2 and prediction errors. These conclusions can be verified by plotting a Taylor diagram (Fig.7), which summarizes multiple aspects of model performance, such as the agreement and variance between observed and predicted values. Taylor diagrams interpretation relies on the relationships between explained variance and bias (from observed and modelled data). Taylor diagram shows that SVM is more distant than the other implemented approaches. Also, the RF method is closer to the observed value, followed by QRF. Although, there is no significant difference was observed between the RF and QRF algorithms.

Table 3. Performance of three different DSM modeling methods to predict soil salinity

Model	MAE	RMSE	R^2
RF	1.722	3.367	0.701
QRF	1.724	3.548	0.695
SVM	2.955	3.628	0.505

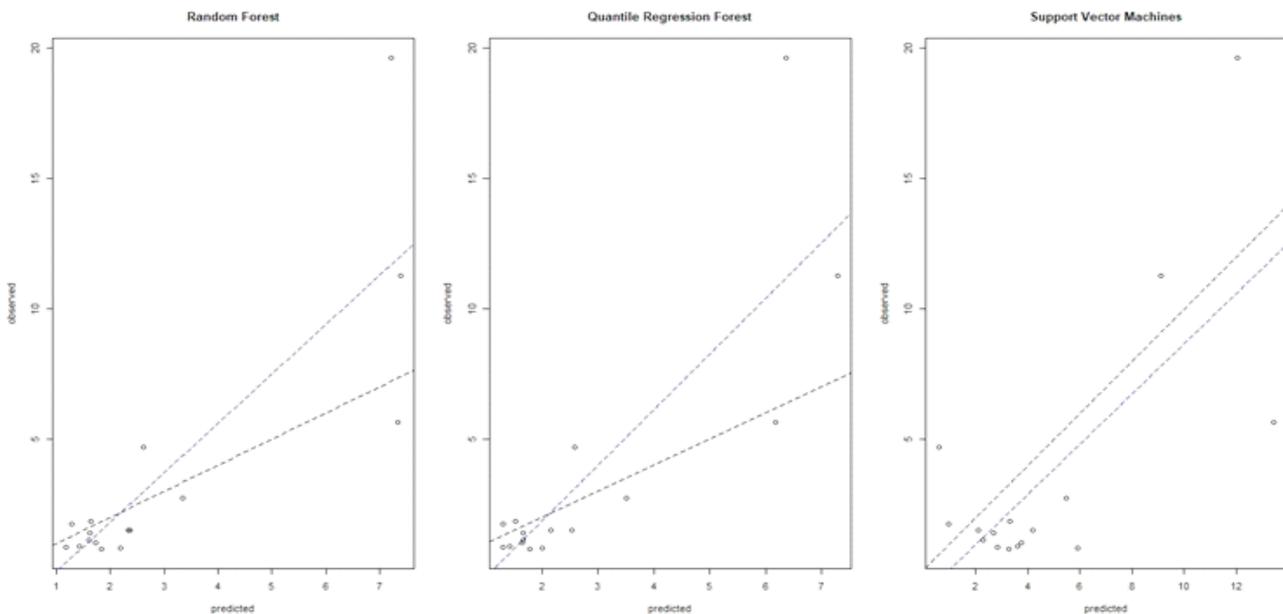


Fig. 6. Scatter plots of predicted against observed values in different DSM models

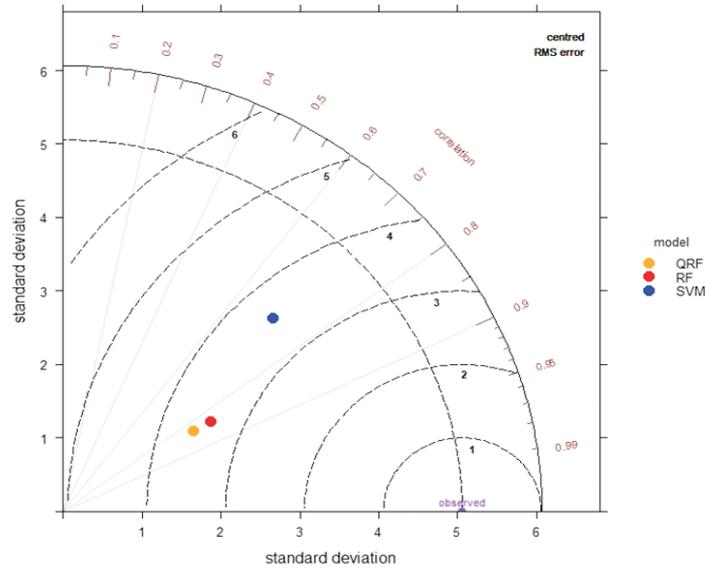


Fig. 7. Taylor diagram used in the evaluation of the three selected DSM models

Spatial prediction of EC_e

Fig. 8 shows the map of EC_e obtained by the best model. Wide variation in distribution of salinity was observed in the study area. The spatial patterns of EC_e in all models are sensible with large values in the Anupgarh

branch canal side of the study area. The spatial patterns of EC_e showed high levels in Anupgarh branch of canal area, and low levels in the main and Bikaner canal area. The variation of salinity in the 5 km vicinity of IGNP canal system, especially with regard to the quantity and distribution of salts in each branch canal, differed

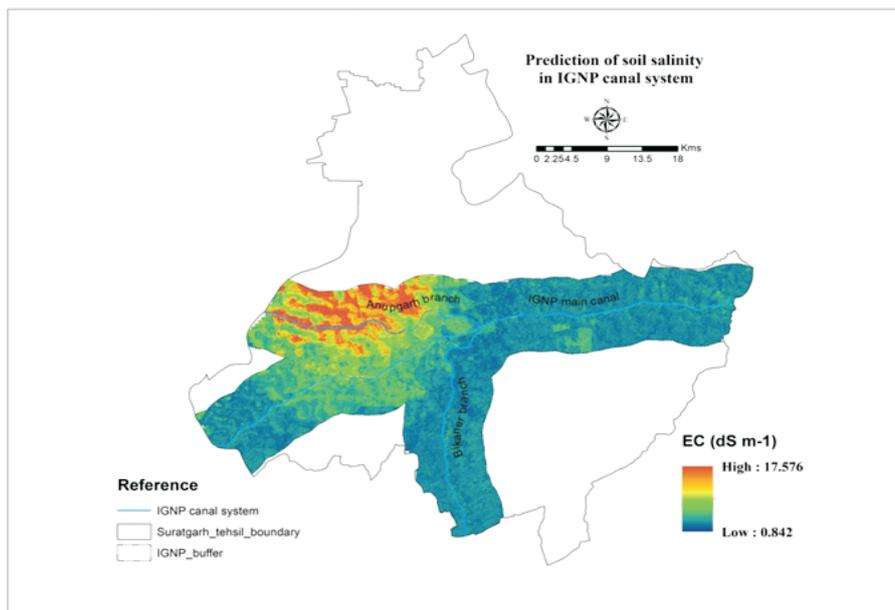


Fig. 8. Spatial distribution of soil salinity (EC_e) using the RF model in the IGNU canal system.

considerably. Hydrology was the main factor responsible for the difference in accumulation and distribution of soluble salts in soil profiles. Continuous water supply, intense evaporation, and the existence of potential sources of salts in the profile or input of salts with water are main cause of development of secondary salinization in Anupgarh branch of canal area. Presently, the aquifer system in the Anupgarh branch area is made up of a complex arrangement of layers of sand and clays (Khan *et al.* 2003). It has frequent lenses of silt, clay and kankar and occasional gravel horizons. Presence of a hard pan is a hydrological barrier that does not allow the percolated water to infiltrate deeper. The presence of impervious layers accompanied by absence of surface drainage outlet is a major factor in rising water table and subsequent waterlogging in lowland, which is responsible for secondary salinization.

Conclusion

This study shows that DSM algorithms can be successfully used for mapping salinity and their uncertainty at a large scale in the IGNP command area of India where there is a wide range in land use and terrain attributes. Among the DSM algorithms, RF model showed the best performance in predicting EC at the regional level, achieving the largest accuracy compared to the other algorithms tested. However, due to the similar performance of the RF, QRF and SVM models, we suggest that all three models should be calibrated, and then the best results applied for spatial prediction of target soil attributes in other geo-graphical settings. The most important covariates that influence EC_e distribution in the IGNP command area are rainfall, elevation aspect and valley depth. The vulnerability to soil salinity showed high levels in Anupgarh branch of canal, and low levels in the IGNP main and Bikaner canal area. The progressive increase in vulnerability to soil salinity in the Anupgarh branch of IGNP area is attributed to unsustainable agricultural practices and inputs, quality of irrigation water, lack of advanced irrigation technologies and efficient drainage systems, and improper land management. Further more, the model developed in this study provides comprehensive

guidance for the land planners and decision makers to develop amicable strategies, practicing policies and regulations to address the appropriate use of the land and for an improvement of the ecological system.

References

- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J. A. and Zhang, G.L. (2014). Global soil map: toward a fine-resolution global grid of soil properties. *Advances in Agronomy* **125**, 93–134.
- Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* **20**, 273–297.
- Dharumarajan S., Hegde, R., Janani, N. and Singh, S. K. (2019). The need for digital soil mapping in India. *Geoderma Regional* **16**, e00204. <https://doi.org/10.1016/j.geodrs.2019.e00204>.
- Dharumarajan, S., Hegde, R. and Singh, S. K. (2017). Spatial prediction of major soil properties using Random Forest techniques - A case study in semi-arid tropics of South India. *Geoderma Regional* **10C**, 154-162.
- FAO and ITPS (2020). Global Soil Organic Carbon Map V1.5. Technical Report, Rome, FAO. <http://www.fao.org/3/ca7597en/ca7597en.pdf>.
- Gorji, T., Sertel, E. and Tanik, A. (2017). Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecological Indicators* **74**, 384–391.
- Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., de Jesus, J. M., Tamene, L. and Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS One* **10**, 1–26.
- Jaglan, M. S. and Qureshi, M. H. (1996). Irrigation development and its environmental

- consequences in arid regions of India. *Environmental Management* **20**, 323–336.
- Khan, M. A., Moharana, P. C. and Singh, S. K. (2003). Integrated natural resources and environmental impact assessment for sustainable development of Ganganagar district, Rajasthan. Research Report, Division of Natural Resources and Environment, Central Arid Zone Research Institute, Jodhpur, India.
- Kühn, J., Brenning, A., Wehrhan, M., Koszinski, S. and Sommer, M. (2009). Interpretation of electrical conductivity patterns by soil properties and geological maps for precision agriculture. *Precision Agriculture* **10**, 490–507.
- Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modeling*, Springer.
- Mahmoudzadeh, H., Matinfar, H. R., Taghizadeh-Mehrjardi, R. and Kerry, R. (2020). Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Regional* **21**, e00260.
- Manickam, L., Dharumarajan, S., Khandal, S., Hegde, R. (2021). Modeling and Mapping of Salt-Affected Soils through Spectral Indices in Inland Plains of Semi-arid Agro-Ecological Region. *Journal of the Indian Society of Remote Sensing* **49**, 1475–1481.
- Maynard, J. J. and Levi, M. R. (2017). Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. *Geoderma* **285**, 94–109.
- McBratney, A. B., Mendonca-Santos, L. and Minasny, B. (2003). On digital soil mapping. *Geoderma* **117**, 3–52.
- Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research* **7**, 983–999.
- Minasny, B. and Hartemink, A. E. (2011). Predicting soil properties in the tropics. *Earth-Science Reviews* **106**, 52–62.
- Minasny, B., McBratney, A. B., Malone, B. P. and Wheeler, I. (2013). Digital mapping of soil carbon. *Advances in Agronomy* **118**, 1–47.
- Shyampura, R. L., Singh, S. K., Singh, R. S., Jain, B. L. and Gajbhiye, K. S. (2002) Soil series of Rajasthan (NBSS Publ. No. 95). NBSS & LUP, Nagpur, p 364.
- Singh, A. (2015). Soil salinization and waterlogging: a threat to environment and agricultural sustainability. *Ecological Indicators* **57**, 128–130.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F. and Malone, B. P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* **213**, 15–28.
- Wang, Y., Deng, C., Liu, Y., Niu, Z. and Li, Y. (2018). Identifying change in spatial accumulation of soil salinity in an inland river watershed, China. *Science of the Total Environment* **621**, 177–185.