GLOBAL SYMPOSIUM ON
SALT-AFFECTED SOILS

20 - 22
October, 2021
Virtual meeting

Spatial Predictability of Salinity Hazard with Machine Learning Algorithms and Digital Data in the Irrigation Plain

Fuat KAYA

INTRODUCTION

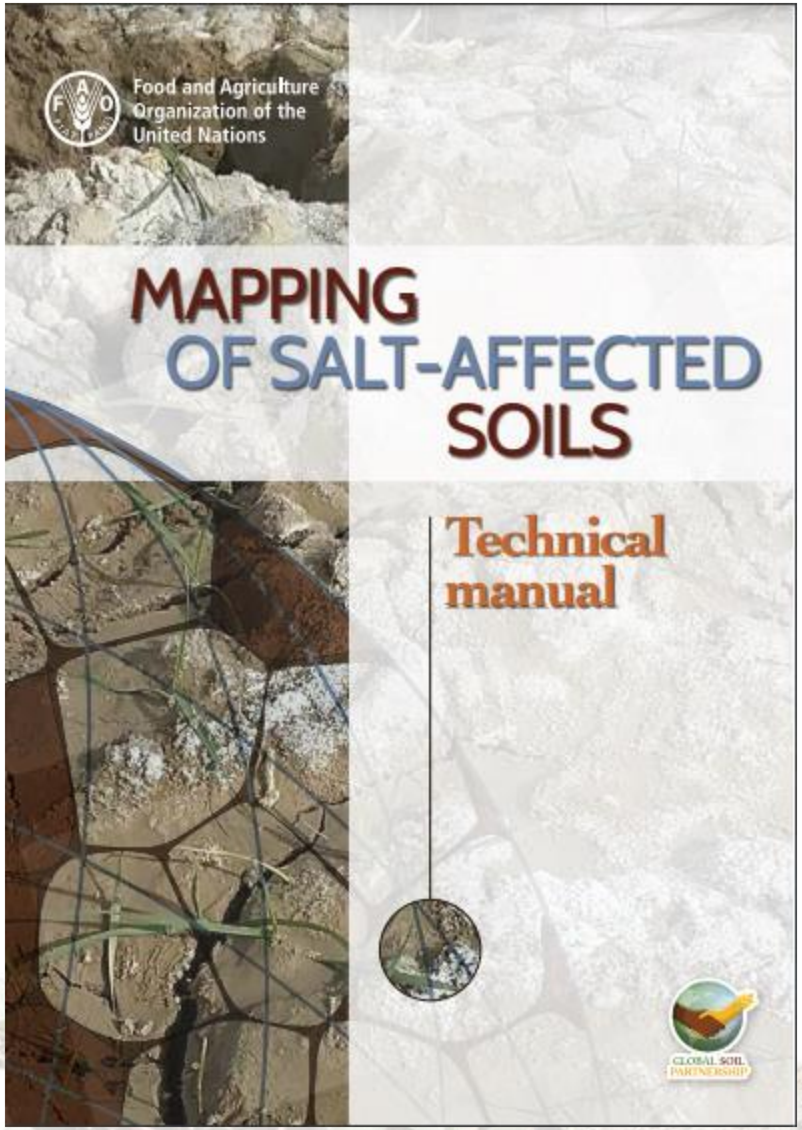
- In the irrigation area in semi-arid regions, spatially detection studies are important marker for accurate monitoring of soil salinization.
- Soil scientists are interested in spatially accurate prediction of the Electrical Conductivity (EC) of soils, which are being greatly interfered with by humans.
- Irrigation applications and additional organic and chemical fertilizer applications are applied to the soil in most of the arid and semi-arid areas. As a result of irrigation, capillarity, which is the soil dynamic system, tends to increase soil salinity on the surface.

INTRODUCTION

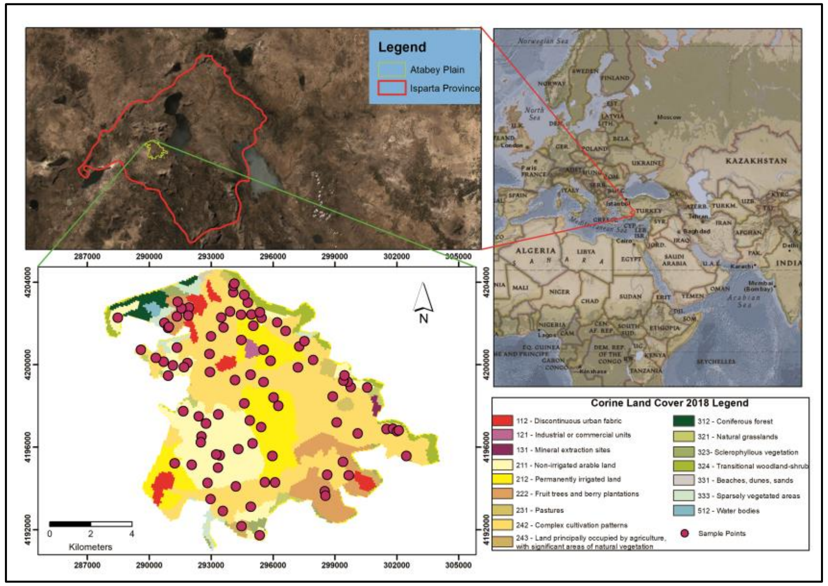
- Machine learning-based modeling approaches in agricultural systems are used to suggest decision/support systems.

- In digital soil mapping, the usability of soil electrical conductivity (EC) information obtained from geostatistical approaches is limited in areas where there are socio-ecologically heterogeneous farms, that is, different land uses. However, the spatial location of training and test points is often neglected in the process of mapping studies with machine learning algorithms currently used to generate spatial predictions. As a result, only data-driven modeling results overfitting problems may be encountered.

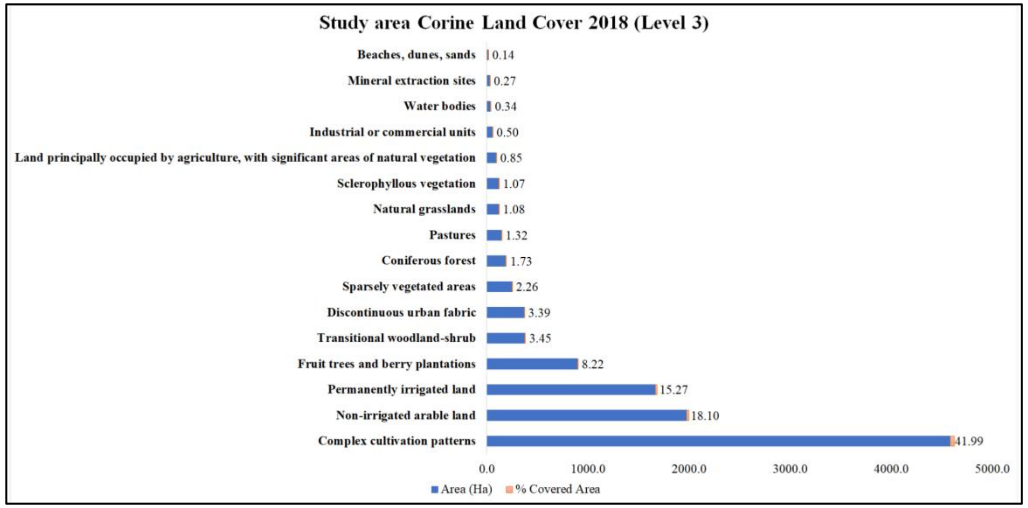
- We immediately need proven methods to measure and map the spatial distribution of soil salinity to manage the processes that drive soil salt transport into the root zone.

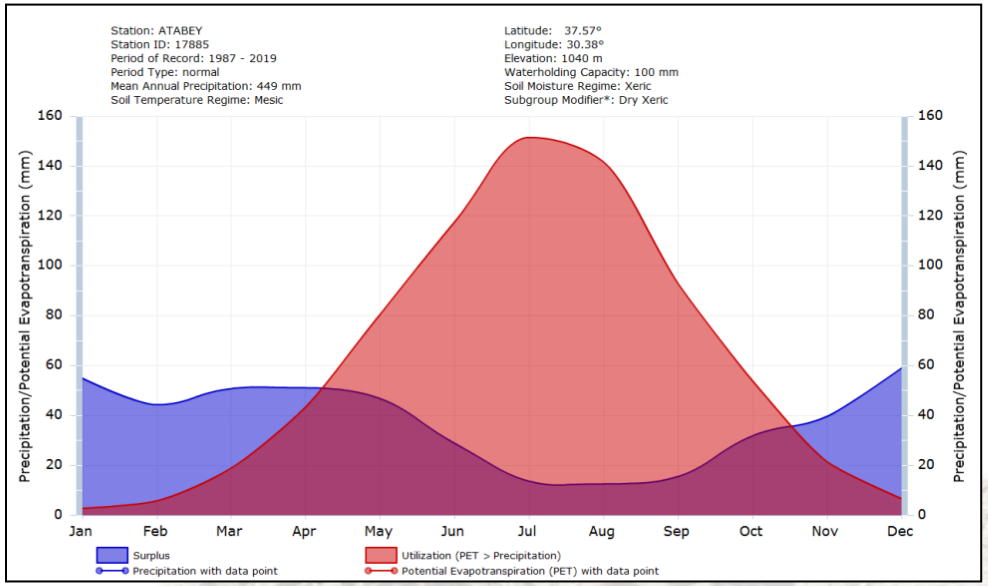


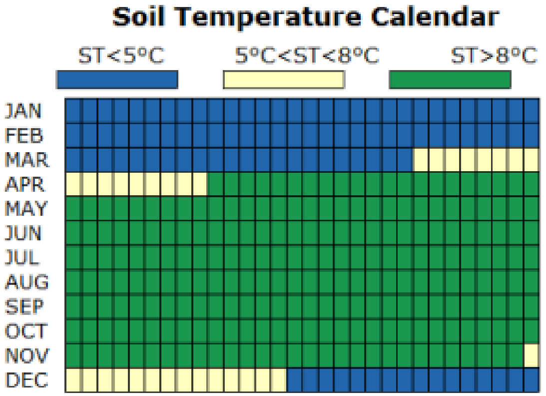
MATERIAL

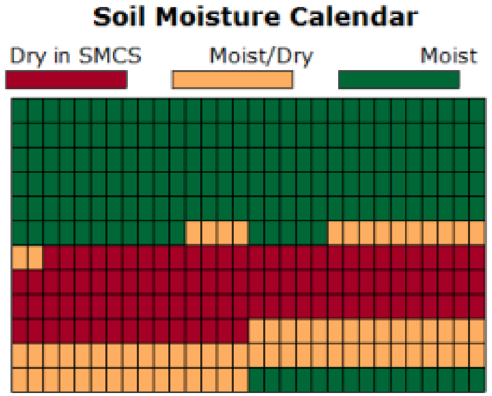


Irrıgation Modernization Project (P158418) Atabey Plain Irrigation Rehabilitation Project





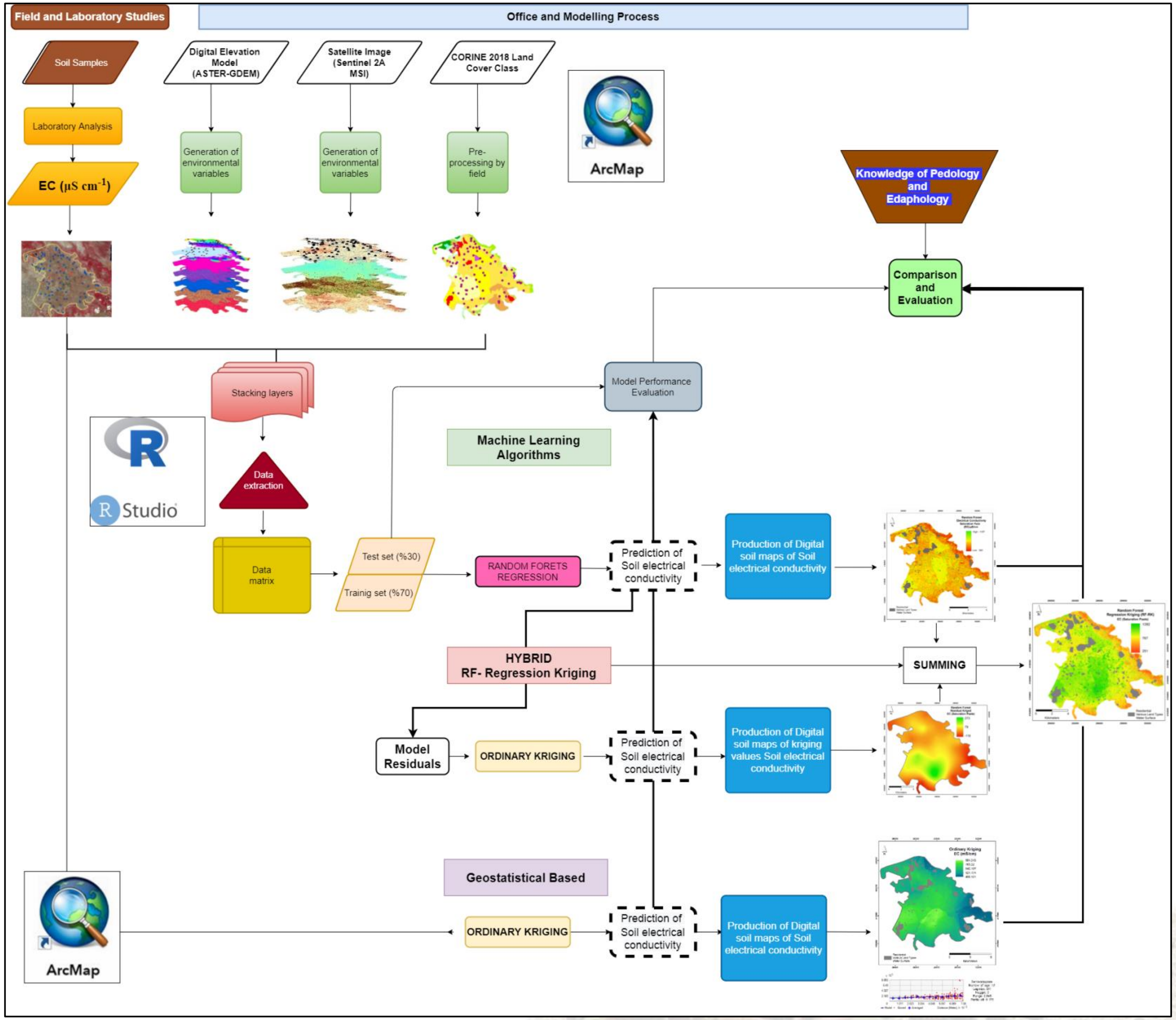




Newhall simulation model jNSM

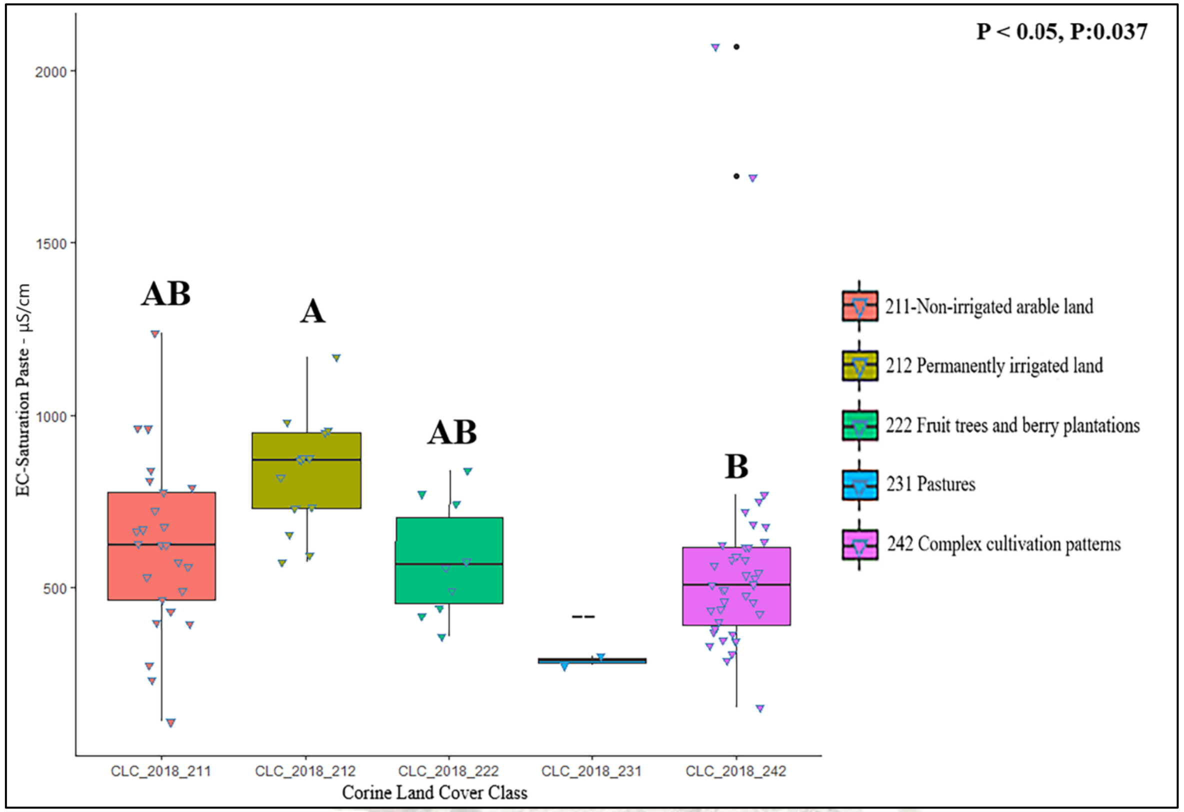
METHODOLOGY

The study was carried out in an agricultural plain where irrigation activities have been carried out for a quarter of a century. EC (EC-µS cm^{-1}) was determined in the saturation paste for 91 samples taken from the field. Environmental variables were generated from Sentinel 2A-MSI satellite, Digital Elevation Model, and CORINE Land Cover Classes. The data set was divided into 70% training and 30% test set. Relevant packages were used in R Core Environment in data set preparation processes and modeling. Ordinary kriging was applied by controlling the normal distribution of the dependent variable. Also, random forest algorithm spatial modeling was used. In the hybrid (RF- Regression Kriging) approach, explanatory variation is estimated by RF algorithms and the process is carried out by summing the regression value of EC and the kriging values of model residuals in non-sampled locations. Root mean square error (RMSE) values were used as model accuracy criteria.



RESULTS-DESCRIPTIVE STATISTICS

EC-µS cm^{-1}	ALL DATA SET	TRAINING	TESTING
Count	91	63	28
Mean	612.1	648.2	530.9
StDev	288.5	302.1	240.9
Variance	83258.3	91288.1	58022.2
CoefVar	47.14	46.61	45.37
Min	110	274	110
Q1	430	434	333.8
Median	574	595	511
Q3	732	769	660.3
Max	2068	2068	1169
Skewness	2.08	2.41	0.57
Kurtosis	7.94	8.65	0.65



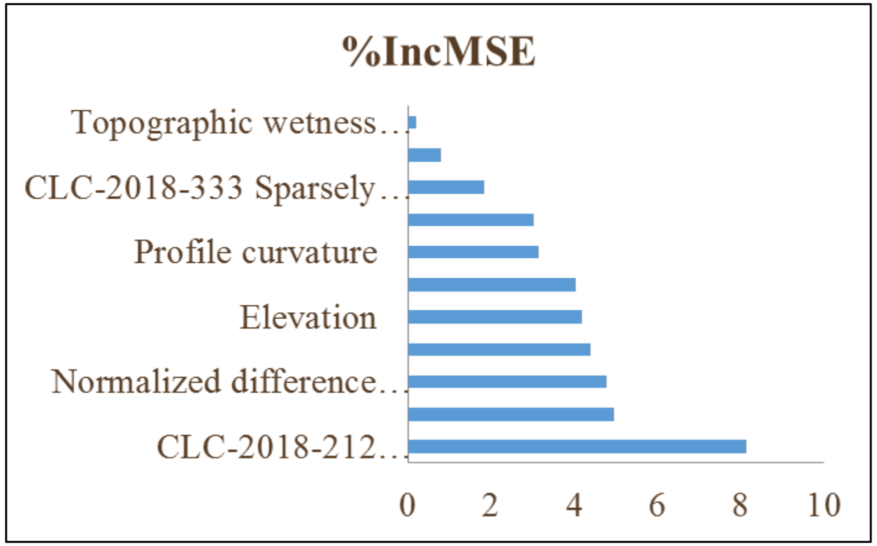
- Comparison of EC-µS cm^{-1} of soil samples according to land cover classes. Boxplots and "CLCC_2018" illustrate the distributions and comparisons of means of EC-µS cm^{-1} across land cover class categories. Within each panel, the means of groups not sharing letters differ significantly at $p < 0.05$ in Tukey's HSD tests.

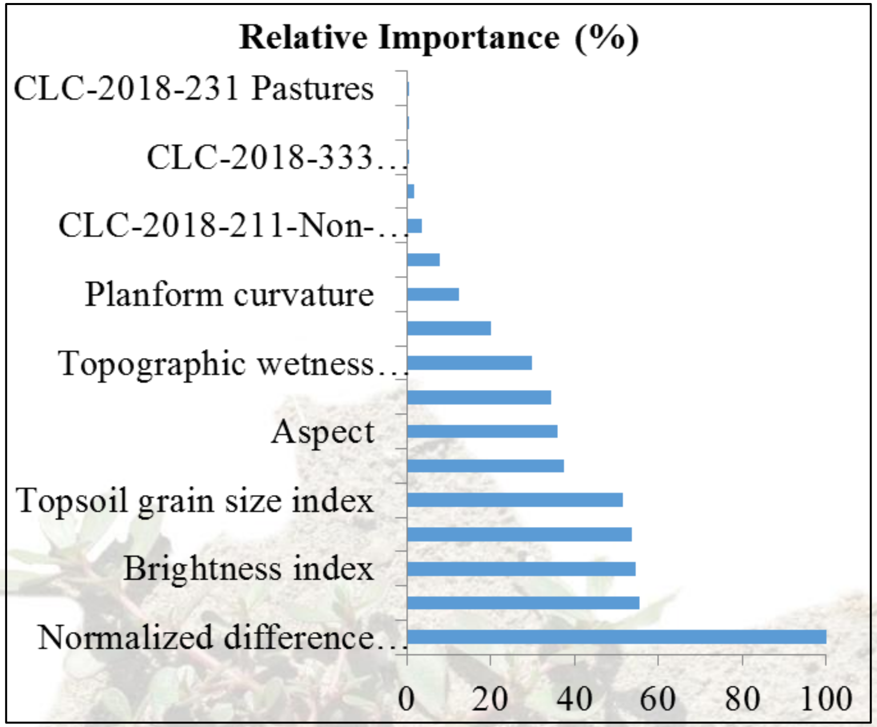
RESULTS-MODEL PERFORMANCES CRITERIA

	Geostatistical Model	
	Cross Validation	
EC-μS cm^{-1}	Model	RMSE
	Ordinary Kriging	270.8

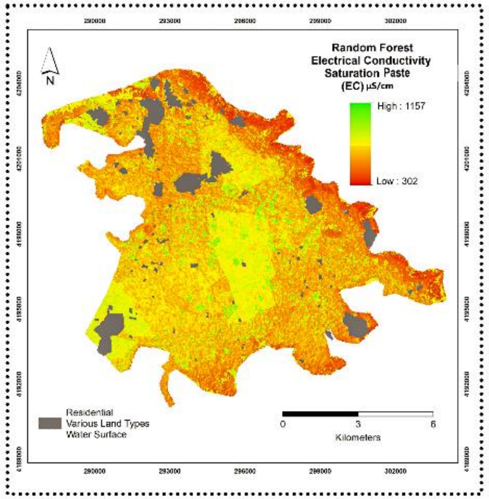
	Machine Learning Model						
		Training			Testing		
EC-μS cm^{-1}	RF	R^2	Lin's concordance	RMSE	R^2	Lin's concordance	RMSE
		0.83	0.87	102.43	0.24	0.28	314.00

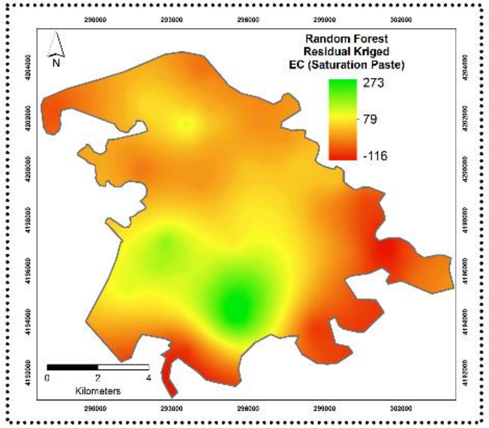
	Hybrid Model						
		Training			Testing		
EC-μS cm^{-1}	RF -RK	R^2	Lin's concordance	RMSE	R^2	Lin's concordance	RMSE
		1	0.98	5.88E-14	1	0.964	9.55E-14

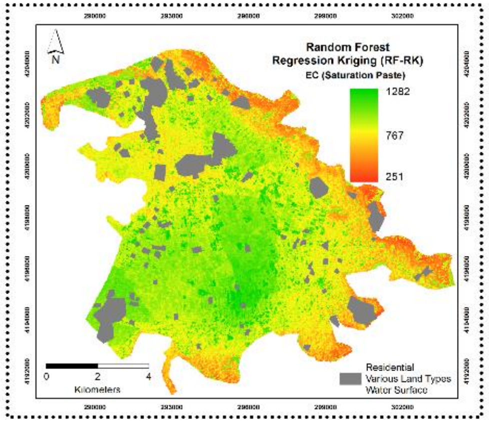


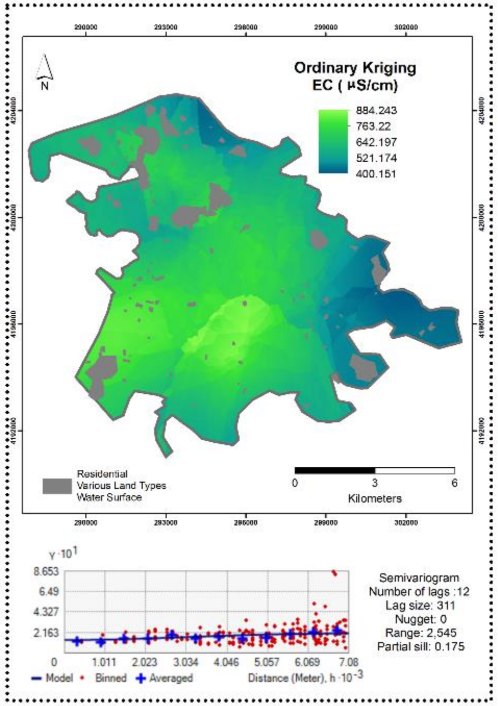


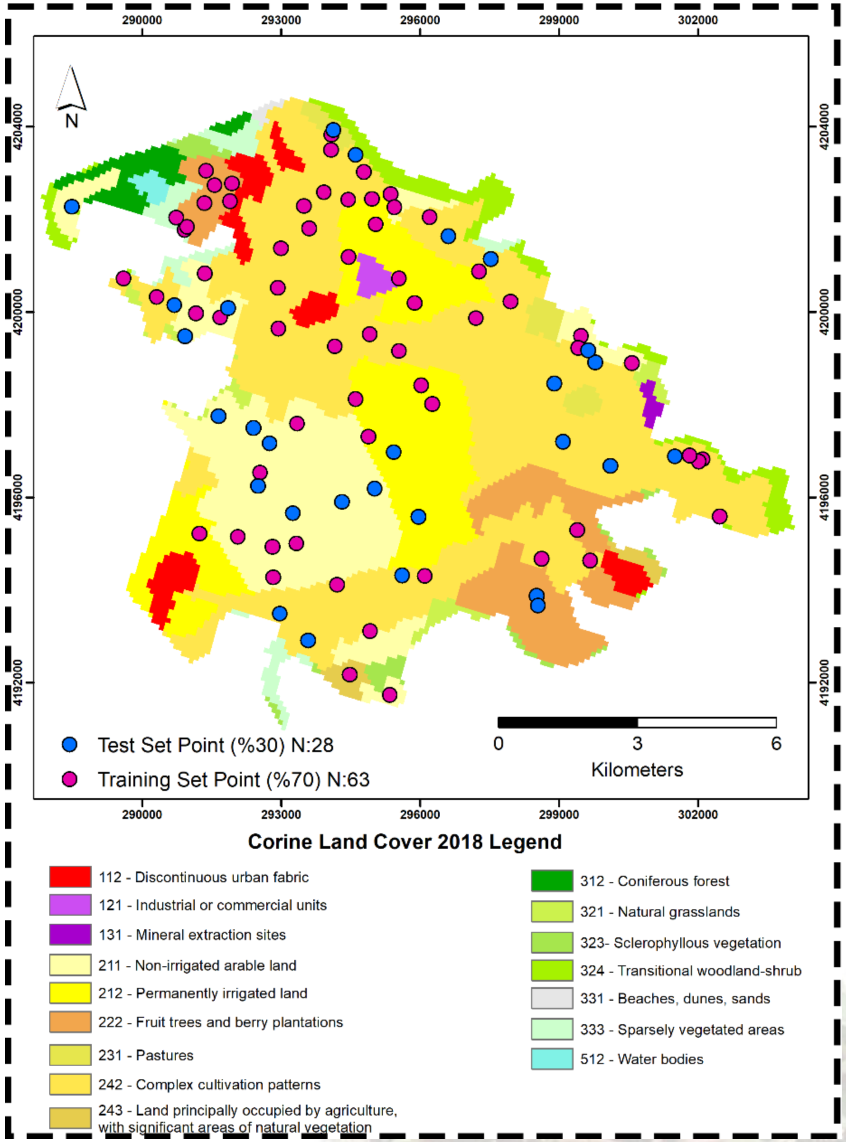
RESULTS-PREDICTIVE MAPPING











DISCUSSION

- The approach of machine learning-based modeling was given relatively accurate results compared to the geostatistical-based modeling, furthermore, the hybrid modeling technique obtained more accurate modeling results than both approaches.

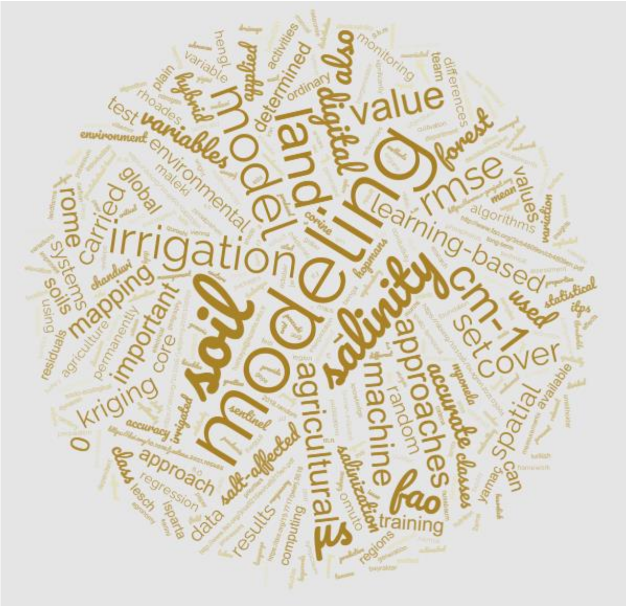
- In machine learning-based modeling approaches, the location of sample points can also be neglected. Land cover class and NDVI value were found to be important environmental variables in the random forest model, indicating that agricultural activities carried out on the land are also important for salinity risk.

CONCLUSIONS

- Machine learning-based modeling approaches using land cover classes as environmental variables may be preferred to mapping soil salinity using a purely geostatistical method.
- In hybrid modeling approaches, the spatial relationship present in model residuals significantly improves model accuracy. It can provide more accurate insights into salinity management and monitoring from digital maps produced.

GLOBAL SYMPOSIUM ON SALT-AFFECTED SOILS

20 - 22
October, 2021
Virtual meeting



Thanks to all participants..