GLOBAL SYMPOSIUM ON

SALT-AFFECTED SOILS

20 - 22
October, 2021
Virtual meeting

Prediction of soil salinity using a Random Forest-based model between 2000 and 2016

A case study in the Great Hungarian Plain

Ghada Sahbeni

Content

Overview

Materials & Methods

Results

Conclusion

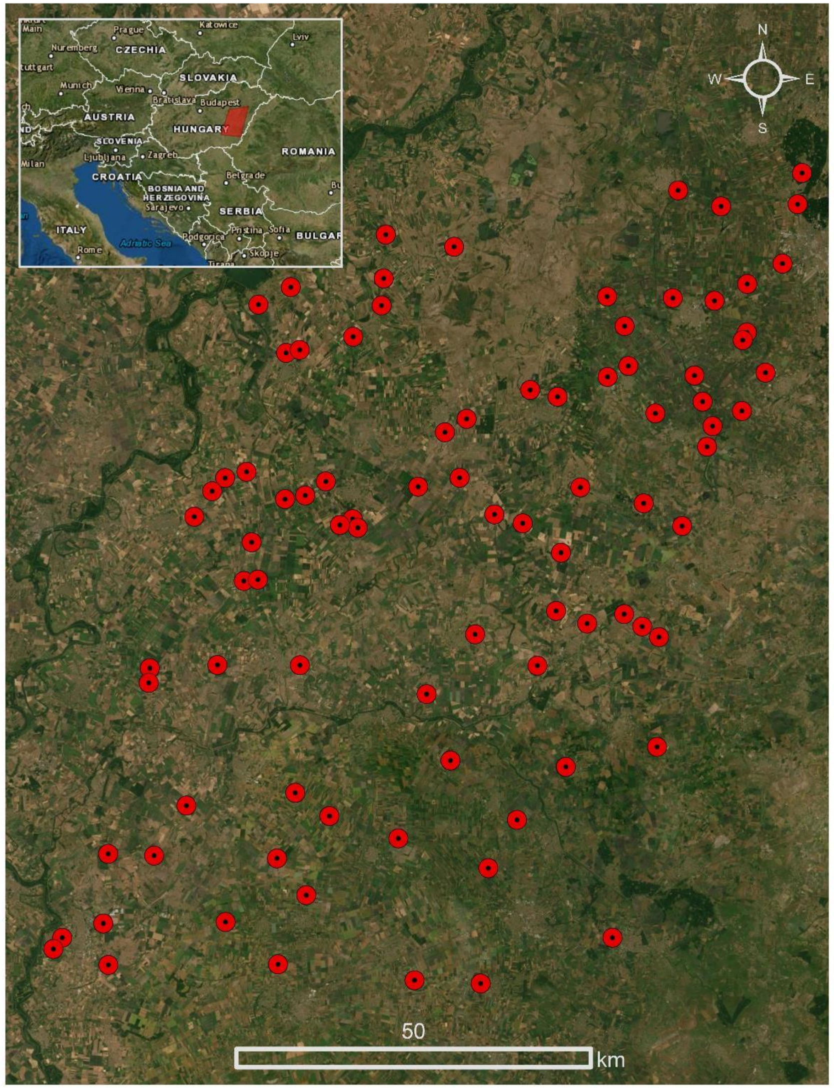
Overview

- Soil salinization threats sustainable agriculture and food security (Shahid et al., 2018).
- According to FAO/UNESCO soil map of the world (1970-1980), 831 million hectares are affected by salinization, with 397 million hectares as saline soils.
- 13 % of the Hungarian territory are salt-affected soils (Tóth, 2009).
- Hydraulic properties of soil and groundwater level influence salt-affected soils genesis in Hungary (Schofield et al., 2001).
- Salinization mapping: A challenging task due to its dynamic nature and fast-expansion.
- Development of adequate restoration plans to preserve soil quality and maintain agricultural productivity.
- Objective: Predict soil salinity using optical and radar data coupled with Random Forest-based models between 2000 and 2016.

Materials & Methods

Study area & Field data

- Eighty-nine samples within soil's upper layer (30 cm) in the Hungarian Soil Monitoring System framework
- Conducted in dry season to detect salts' spectral properties during their accumulation (Szabó and Pirkó, 2017)
- Hungarian standard MSZ-08-0206/2-1978 to measure soil salinity from saturated paste (MSZ 08-0206-2, 1978)
- Meadow chernozems and humic sandy soils domination in the landscape with an expanse of agricultural land cover (Pásztor et al., 2018)
- Main river: Tisza
- Average elevation = 89 meters
- **A moderately warm-dry climate:** a mean yearly precipitation of 560 mm and a mean yearly evaporation of 900 mm (Hungarian Meteorological Service, 2021)



Service Layer Credits: Esri, HERE, Garmin, (c) OpenStreetMap contributors, and the GIS user community
Source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community

Figure 1: Sampling sites location

Remotely sensed data

- Nine Landsat images acquired in 2000, 2008 and 2016 between May and September
- Two ERS-1/2 Medium Resolution Image and one Sentinel-1 SAR Ground Range Detected (GRD) products

Optical data processing

1. Data Acquisition
2. Radiometric Calibration
3. Atmospheric Correction (FLAASH)
4. Spectral Indices
5. PCA

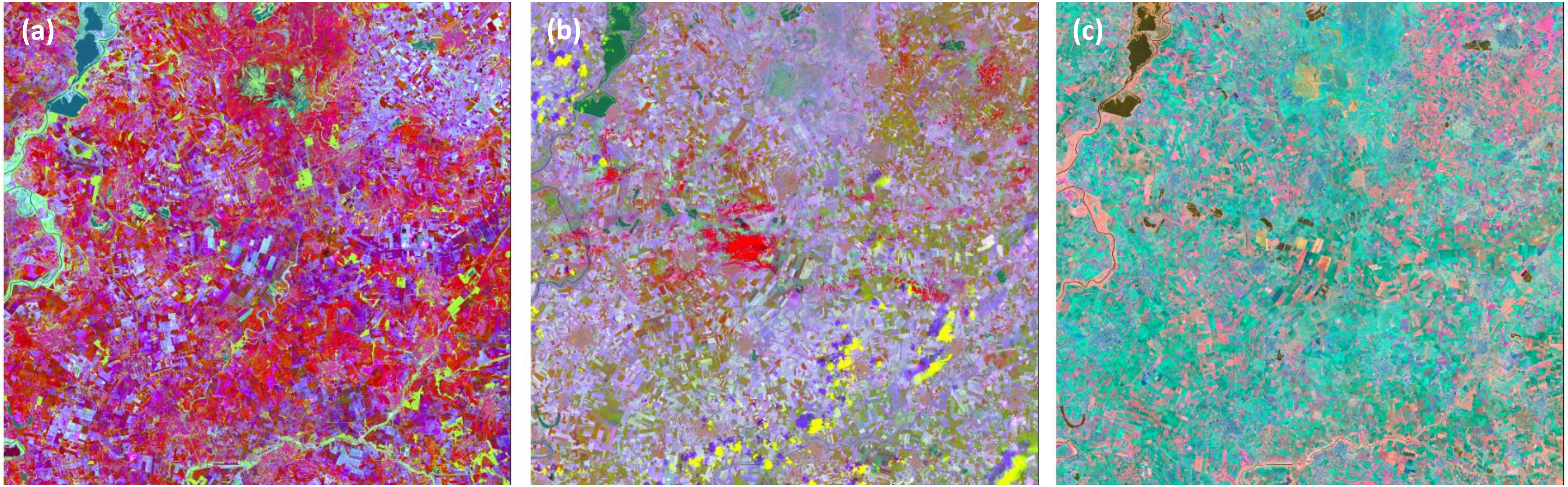
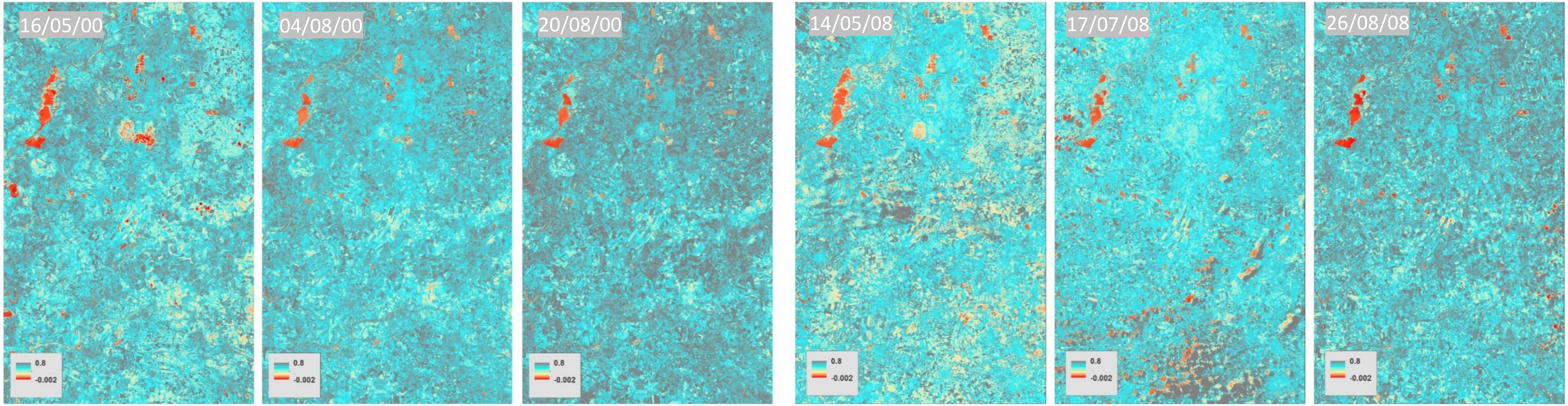
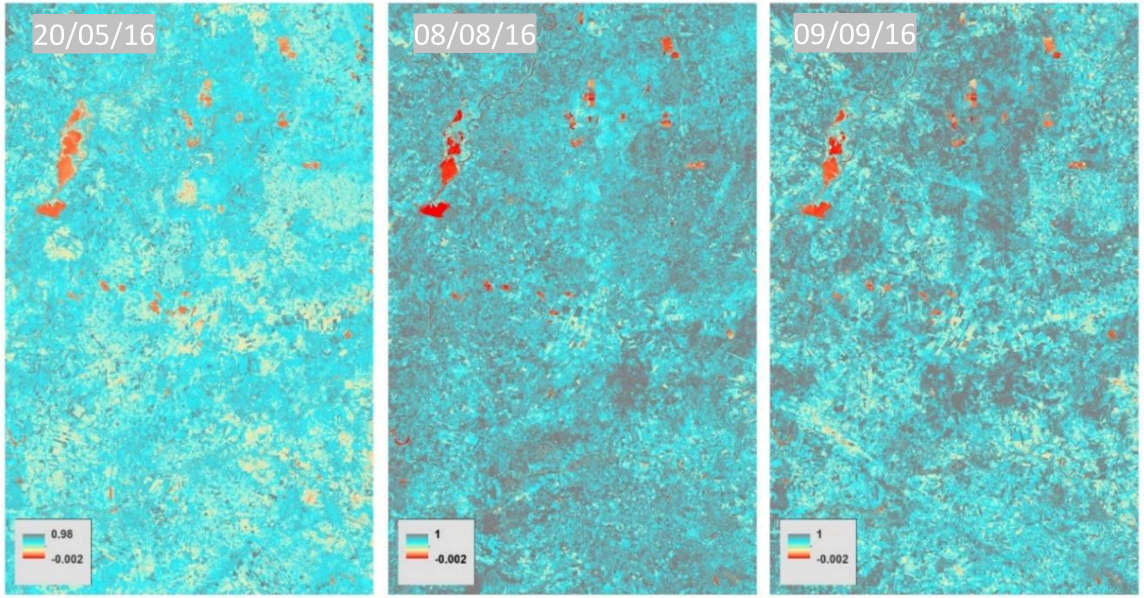


Figure 2: PCA visualization in RGB color composite using stacked Landsat images, (a) PCA 2000, (b) PCA 2008, (c) PCA 2016

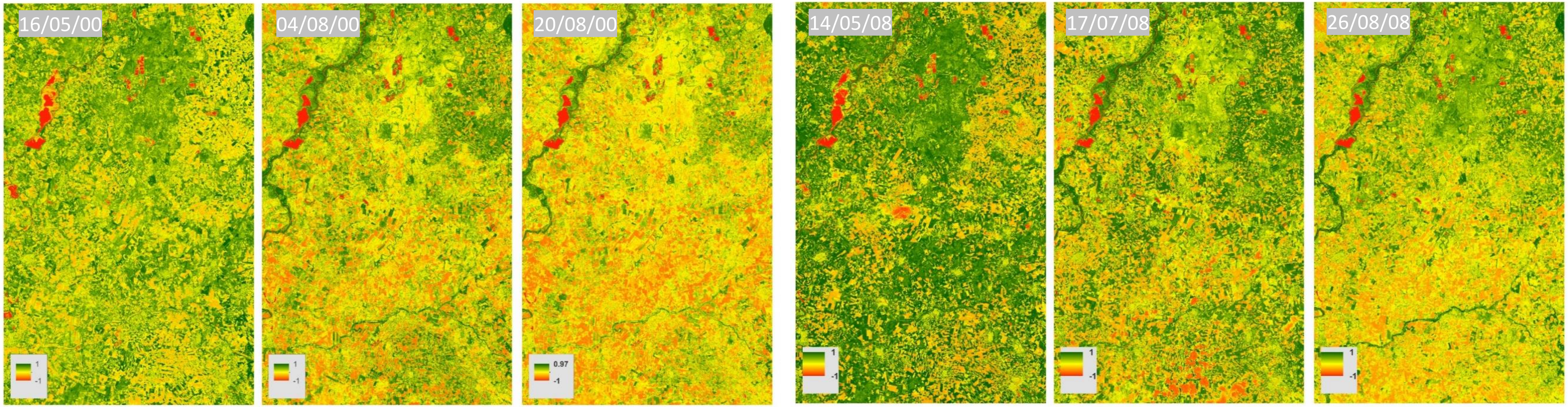


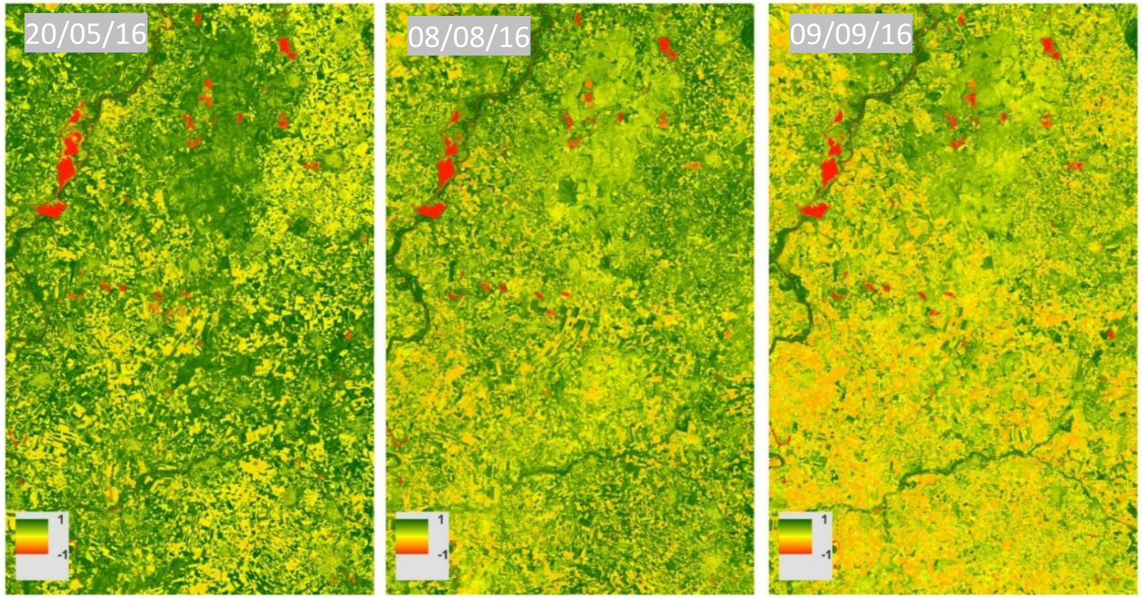


albedo

albedo = ((0.356* B) + (0.130 * R) + (0.373 *NIR) + (0.085* $SWIR_1$) + (0.072* $SWIR_2$) – 0.0018) /1.016

(Silva et al., 2016)

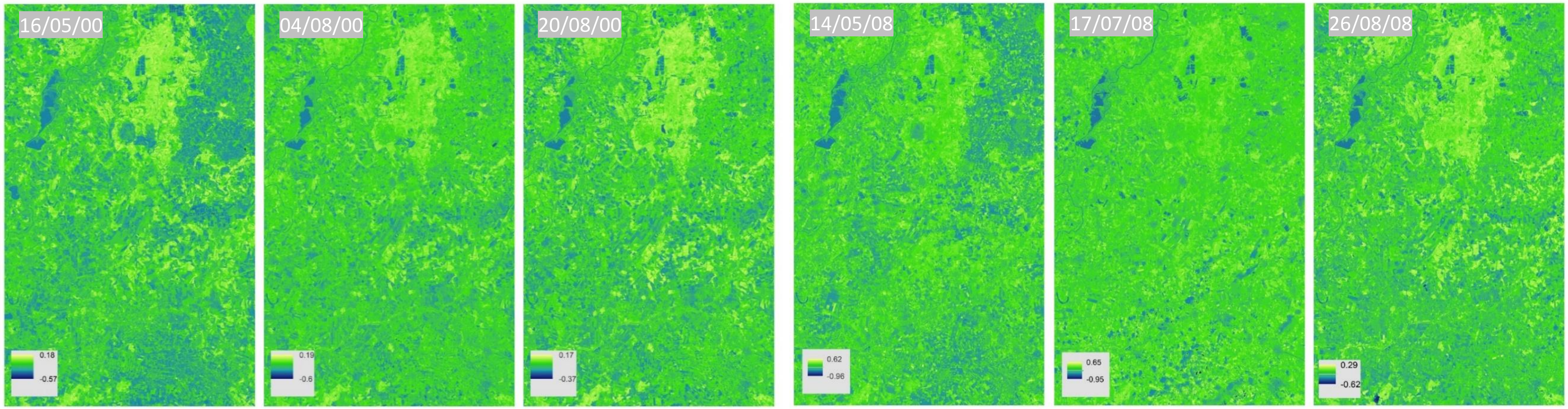


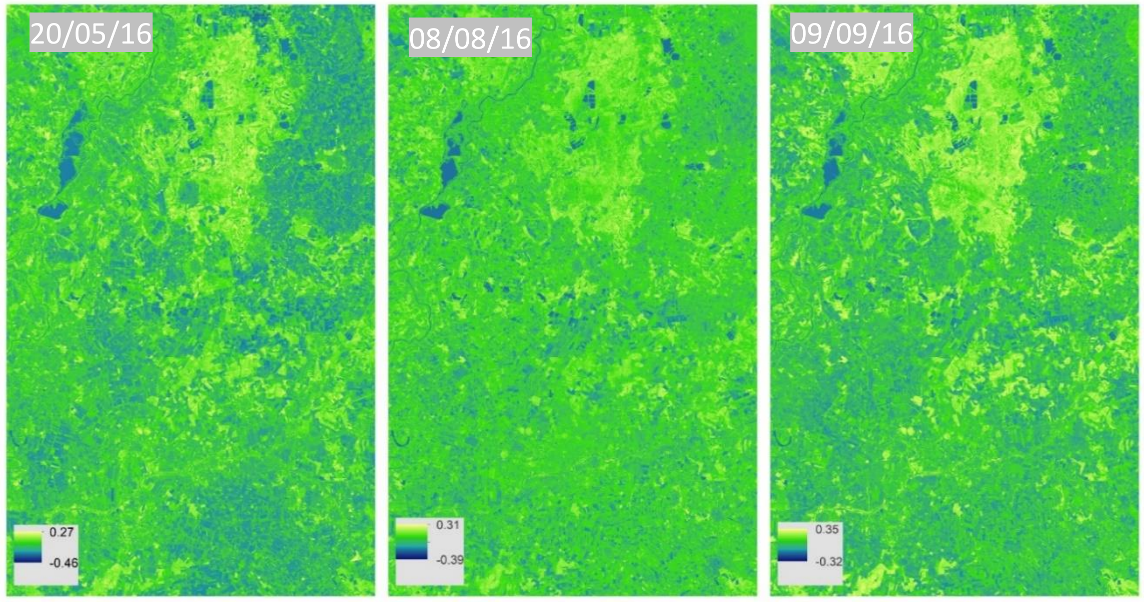


Normalized Difference Vegetation Index

NDVI = (NIR − R) / (NIR + R)

(Rouse et al., 1974)





Soil Salinity and Sodicity Index

$$SSSI = SWIR_1 - SWIR_2$$

(Bannari et al., 2008)

Radar data processing

Apply orbit file → Radiometric Calibration → Speckle filtering → Range Doppler terrain correction → Conversion to dB scale → Coregistration → SAR features extraction

Figure 3: Workflow of radar data processing for ERS-1/2 and Sentinel-1 SAR sensors

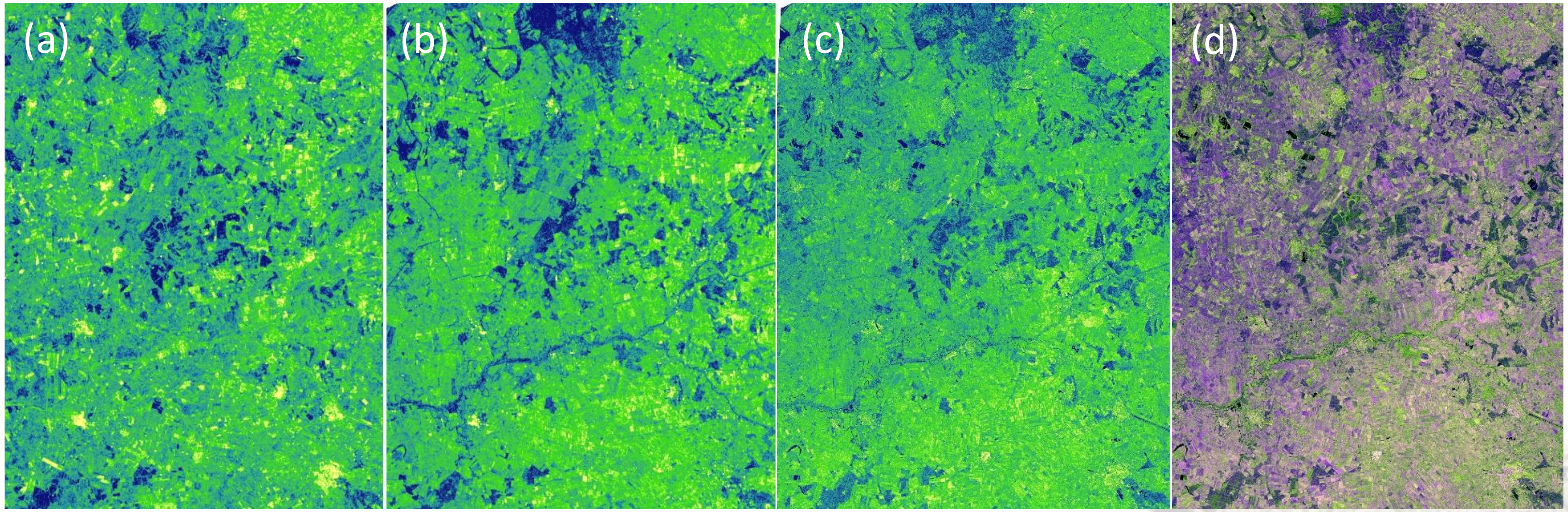


Figure 4: Radar data visualization, (a) VV 2000, (b) VV 2008, (c) VV 2016, (d) RGB 2016

Random Forest

A supervised learning algorithm that uses **ensemble learning** method for regression.

Ensemble learning: Combination of predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

Able to work with missing data, obtain non-linear relationships between independent and dependent features.

Slow to create predictions, unable to extrapolate outside of unseen data, **a predictive not descriptive tool.**

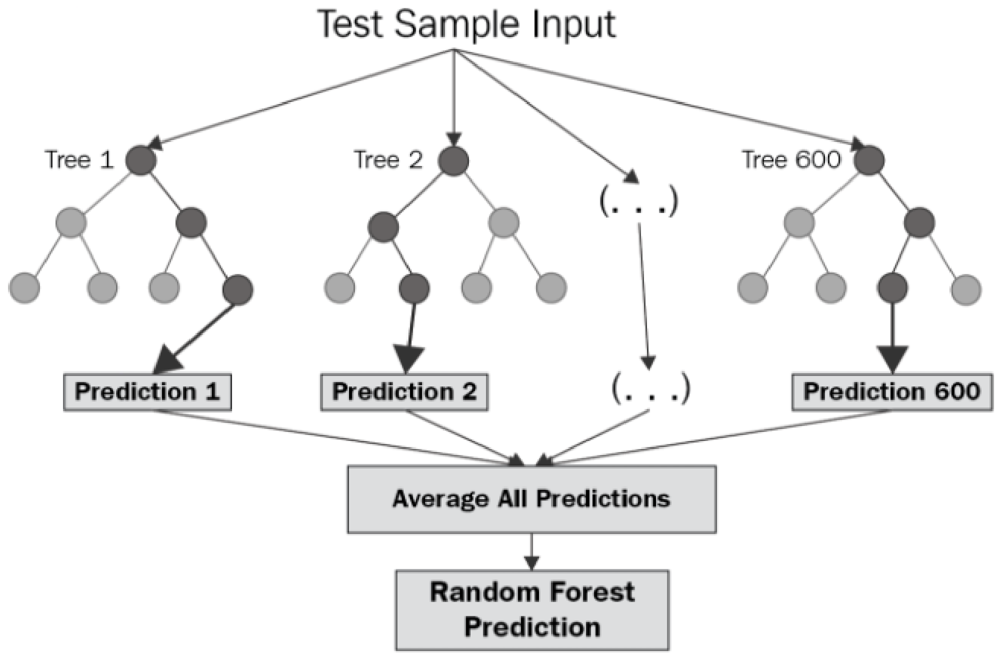


Figure 5: Basic concept of Random Forest
Source: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Random Forest application

- Identify dependent (y) and independent variables (x)
- Split the dataset into training set and test set (80% and 20%)
- Training Random Forest regression model on dataset, hyperparameters calibration (n_estimators = number of trees, m_try = number of variables randomly sampled as candidates at each split), and features selection using recursive feature elimination (RFE) algorithm.
- Application of calibrated model on test set
- Performance assessment using RMSE and correlation coefficient (r)

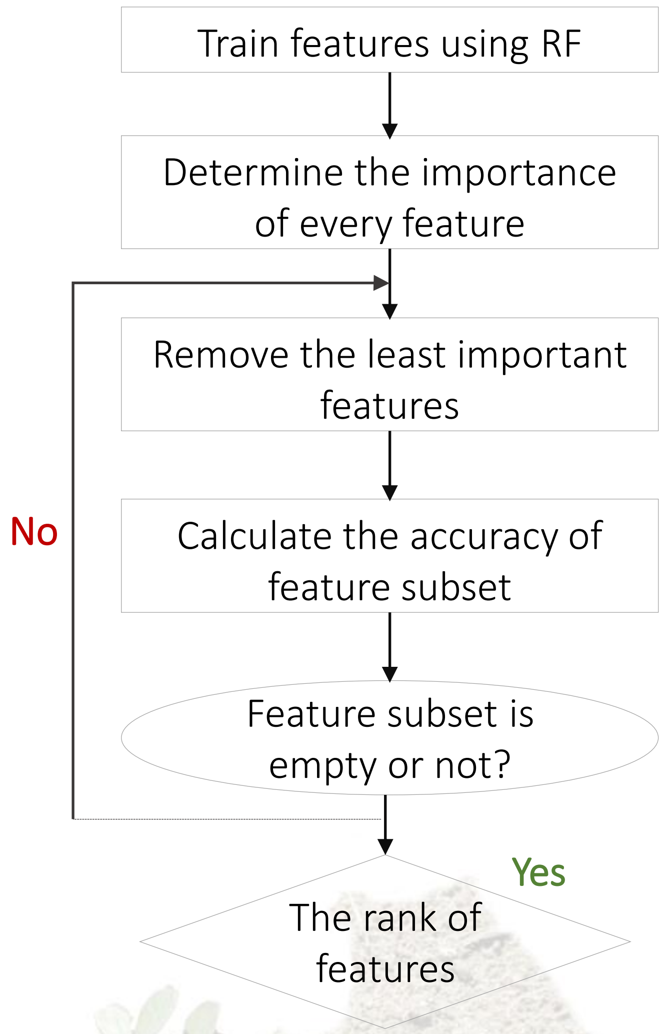


Figure 6: Recursive Feature Elimination (RFE) for variables selection (Qi et al., 2018)

Results

Descriptive analysis

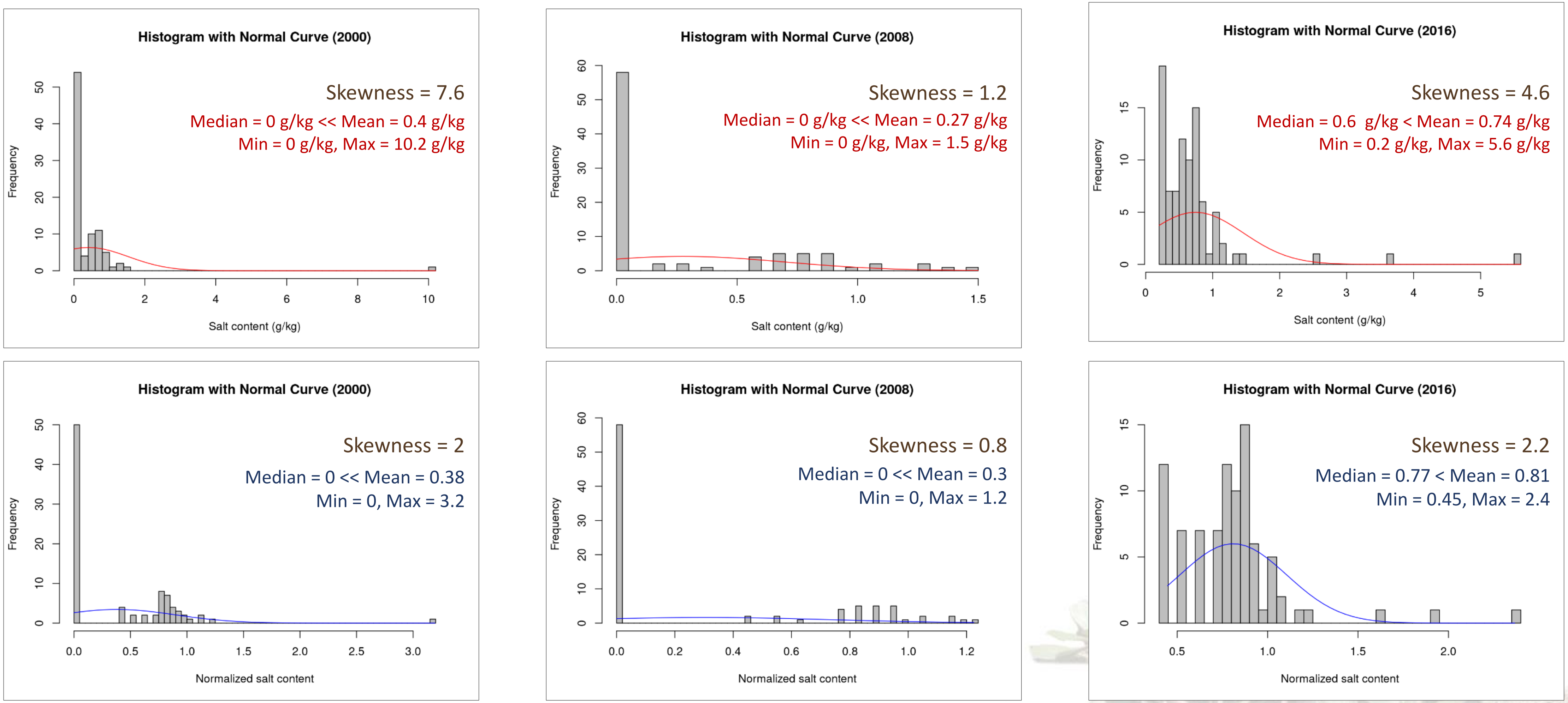


Figure 7: Field data distribution and normalization using square root transformation

Hyperparameters calibration

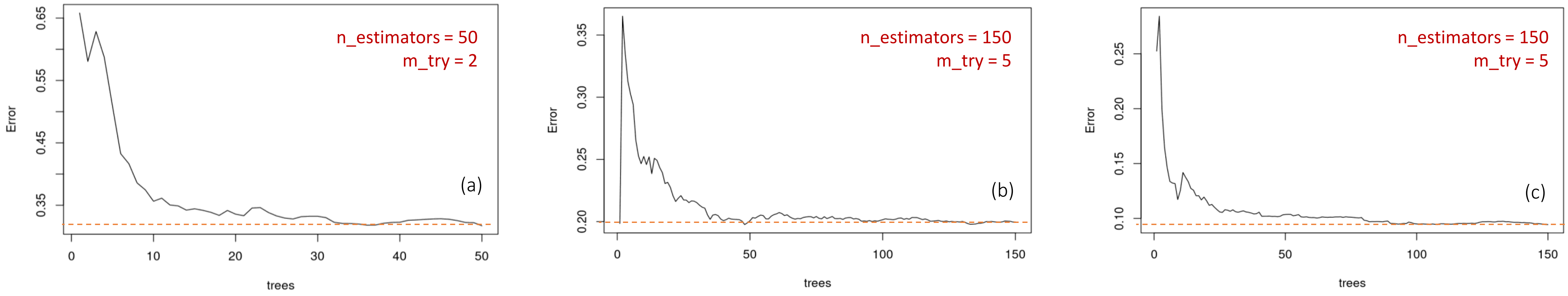


Figure 8: Random Forest based-models, (a) RF-2000, (b) RF-2008, (c) RF-2016

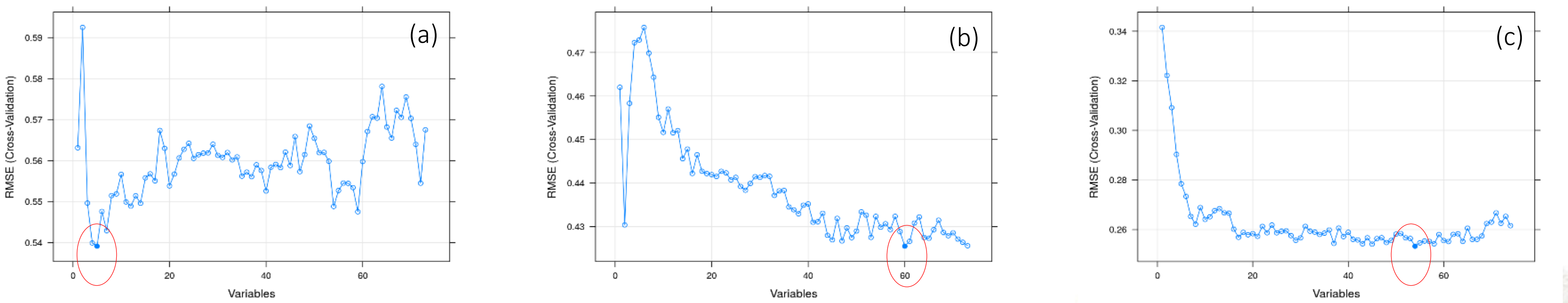


Figure 9: Features selection using RFE, (a) RF-2000, (b) RF-2008, (c) RF-2016

Performance assessment

RF	RMSE Training	RMSE Test	Correlation Coefficient (r)
RF-2000	0.28	0.49	0.73
RF-2008	0.19	0.49	0.8
RF-2016	0.14	0.16	0.81

Table 1: Performance of RF-based models using two statistical metrics

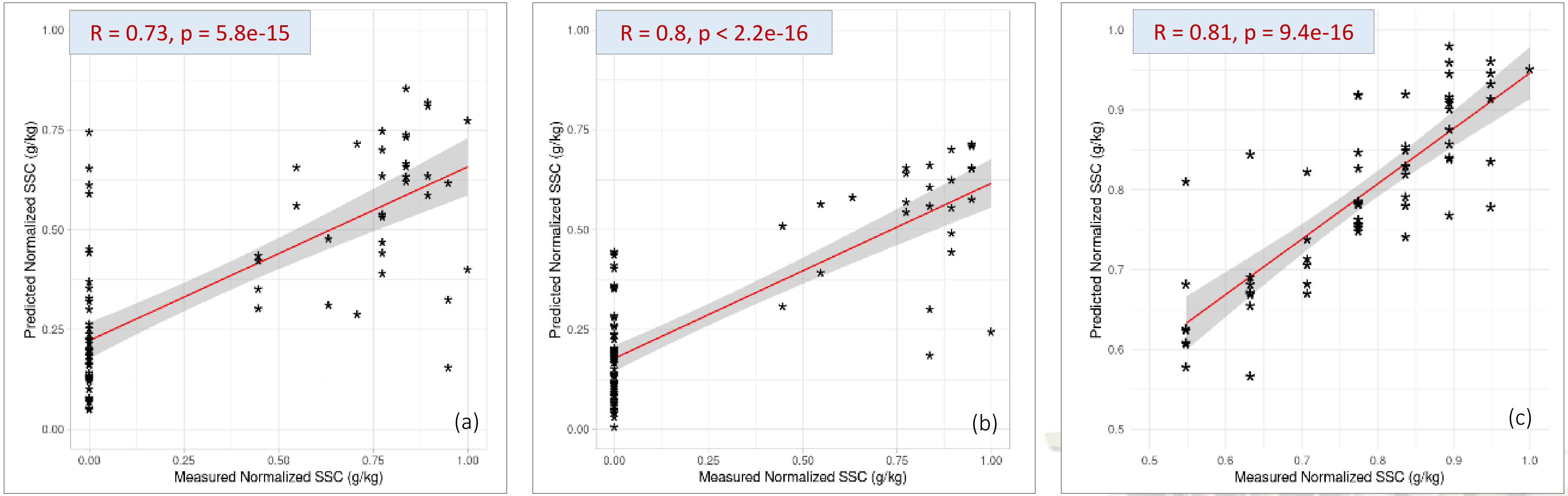


Figure 10: Relationship between measured and estimated normalized salt content values (g/kg), (a) RF-2000, (b) RF-2008, (c) RF-2016

Conclusion

- Random Forest efficiency for soil salinity prediction over time
- An overall significant correlation coefficient (r) of 0.73, 0.8, and 0.81, respectively, for 2000, 2008, and 2016, with an outperformance of 2016-RF model that yielded the lowest RMSE values for training and test sets (0.14 g/kg and 0.16 g/kg)
- Remote sensing tools potential as valuable alternatives in assisting policy-makers to monitor and mitigate salt-affected soils expansion when coupled with robust predictive algorithms
- Dataset size, low variance of samples, spectral noise and atmospheric effect on spectral information retrieved from optical sensors
- The effect of other soil parameters, climatic conditions and groundwater level on salinity variation

GLOBAL SYMPOSIUM ON
SALT-AFFECTED SOILS

20 - 22
October, 2021
Virtual meeting

For questions and potential collaborations, contact me via email:
gsahbeni@caesar.elte.hu
LinkedIn: https://www.linkedin.com/in/ghada-sahbeni/
ResearchGate: https://www.researchgate.net/profile/Ghada-Sahbeni

Acknowledgments

We would like to genuinely thank Prof. L. Pásztor from the Research Institute of Soil Science and Agricultural Chemistry (RISSAC website: https://www.mta-taki.hu/en) for his support to obtain field data.

Field data were acquired in the Hungarian Soil Monitoring System framework between mid-September and mid-October.

Landsat data were downloaded from the United States Geological Survey (USGS). Website: https://earthexplorer.usgs.gov

Sentinel-1 SAR data were downloaded from the Copernicus Open Access Hub. Website: https://scihub.copernicus.eu/dhus/#/home

ERS-1/2 data were downloaded from the European Space Agency SAR Online Dissemination. Website: https://esar-ds.eo.esa.int/oads/access/

References

Bannari A., Guedon A.M., El-Harti A., Cherkaoui F.Z., and El-Ghmari A. (2008). 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,* 39(19-20): 2795. https://doi.org/10.1080/00103620802432717

FAO. Salt-affected soils: Extent of salt-affected soils. Retrieved from: https://www.fao.org/soils-portal/soil-management/manejo-de-suelos-problematicos/suelos-afectados-por-salinidad/more-information-on-salt-affected-soils/es/

Hungarian Meteorological Service. (2021). The climate of Hungary – general characteristics. Retrieved from: https://www.met.hu/en/eghajlat/magyarorszageghajlata/altalanoseghajlatijellemzes/altalanosleiras/

MSZ 08-0206-2. (1978). Evaluation of some chemical properties of the soil. Laboratory tests. (pH value, phenolphtaleine alkalinity expressed in soda, all water-soluble salts, hydrolite, and exchanging acidity.

Pásztor L., Laborczi A., Bakacsi Z., Szabó J., and Illés G. (2018). Compilation of a national soil-type map for Hungary by sequential classification methods. Geoderma 311: 93–108. https://doi.org/10.1016/j.geoderma.2017.04.018

Qi C., Meng Z., Liu X., Jin Q., and Su R. (2018). Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE, *Genes*., 9:301. https://doi.org/10.3390/genes9060301

Rouse J.W., Haas R.H., Schell J.A., and Deering D.W. (1974). Monitoring vegetation systems in the Great Plains, Freden S.C., Mercanti E.P., Becker M., Eds, with ERTS, Third Earth Resources Technology Satellite–1 Symposium, Vol I: Technical Presentations, NASA SP-351, Washington, DC: NASA.

Schofield R., Thomas D., and Kirkby M. (2001). Causal processes of soil salinization in Tunisia, 20 Spain, and Hungary. *Land Degradation & Development,* 12: 163-181. https://doi.org/10.1002/ldr.446

Shahid S.A., Zaman M., and Heng L. (2018). Soil Salinity: Historical Perspectives and a World Overview of the Problem. In: Guideline for Salinity Assessment, Mitigation, and Adaptation Using Nuclear and Related Techniques. Springer, Cham. https://doi.org/10.1007/978-3-319-6 96190-32

Silva B.B., Braga A.C., Braga C.C. Oliveira de LMM, L.M.M., Montenegro S.M.G.L., and 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

Szabó J., and Pirkó B. (2017). The soil information and monitoring system (TIM). Retrieved from: http://eagri.cz/public/web/file/519224/27BelaPirko.pdf.

Tóth T. (2009). Monitoring, predicting and quantifying soil salinity, sodicity and alkalinity in Hungary at different scales: Past experiences, current achievements and outlook with special regard to European Union initiatives, Advances in the assessment and monitoring of salinization and status of biosaline agriculture. Report of an expert consultation held in Dubai, United Arab Emirates, 26-29 November 2007. World Soil Resources Report. No 104. FAO, Rome, 201-229. Retrieved from: https://tibortothsoil.members.iif.hu/abstr/T2009.pdf

Appendix

Spectral Indices

Index	Formula
albedo	$((0.356*B) + (0.130*R) + (0.373*NIR) + (0.085*SWIR_1) + (0.072*SWIR_2) – 0.0018) /1.016$ (Silva et al., 2016)
Differential Vegetation Index DVI	NIR – R (Basso et al., 2000)
Green Normalized Difference Vegetation Index GNDVI	(NIR - G) / (G + NIR) (Wu et al., 2014)
Intensity Index Int_1	$(G + R)/2$ (Bouaziz et al., 2011)
Intensity Index Int_2	$(G + R + NIR)/2$ (Bouaziz et al., 2011)
Normalized Difference Moisture Index NDMI	$(NIR - SWIR_1)/(NIR + SWIR_1)$ (Wilson & Sader, 2002)
Normalized Difference Salinity Index NDSI	(R − NIR) / (R + NIR) (Khan et al., 2005)
Salinity Index 1 SI_1	$\sqrt{G * R}$ (Douaoui et al., 2005)
Salinity Index 2 SI_2	$\sqrt{NIR * R}$ [14]
Salinity Index 3 SI_3	$\sqrt{G^2 + R^2 + NIR^2}$ (Dehni & Lounis, 2012)
Salinity Index 4 SI_4	$\sqrt{R^2 + G^2}$ (Yahiaoui et al., 2015)
Normalized Difference Vegetation Index NDVI	(NIR − R) / (NIR + R) (Rouse et al., 1974)
Soil Adjusted Vegetation Index SAVI	((NIR – R) / (NIR + R + L)) (1 + L) (Huete, 1988)
Enhanced Vegetation Index EVI	$2.5* (NIR − R) / (NIR + C_1*R − C_2*B + L)$ (Huete et al., 2002)
Brightness Index BI	$\sqrt{R^2 + NIR^2}$ (Khan et al., 2005)
Bare Soil Index BSI	(G + NIR) / (G − NIR) (Li et al., 2013)
Normalized Pigment Chlorophyll Ratio Index NPCRI	(R − B) / (R + B) (Merzlyak et al., 1999)
Soil Salinity and Sodicity Index 1 $SSSI_1$	$SWIR_1 − SWIR_2$ (Bannari et al., 2008)
Soil Salinity and Sodicity Index 2 $SSSI_2$	$(SWIR_1*SWIR_2 – SWIR_2*SWIR_2)/ SWIR_1$ (Bannari et al., 2008)
Normalized Difference Water Index NDWI	$(G - NIR)/(G + NIR)$ (Sykas, 2020)
Normalized Difference Salinity Index VSSI	2*G − 5 (R + NIR) (Dehni & Lounis, 2012)

Selected features

RF-2000	RF-2008	RF-2016
bsi3 ndmi3 int12 si12 int13 si42 b3 gdvi2 npcri3 b8 dvi2 ndwi2 ndwi3 gdvi3	b10, npcri2, ndmi2, sssi12 gdvi3, bsi2, gdvi1, ndwi1, ndwi3, b3, savi1, b5, int12, dvi1, b9, evi1, bsi1, b2, gdvi2, ndwi2, ndsi1, sssi21 ndmi1, int11, si11, albedo3, si12, dvi3, si41, b7, int21, ndsi2, b1 savi2, albedo1, si23, ndvi3, albedo2, int22, evi2, ndsi3, si42, int13, ndvi2, ndmi3, dvi2, si33, si13, si32, si22 npcri3, bi1, savi3, si31 bi2, vssi2, vssi1, ndvi1, si43, bsi3	b10, sssi23, albedo1 si21, int21, sssi13 b6, vv16db, int13, vssi1, si43, sssi21, sssi11, si31 bi1, si13, ndwi3, gdvi3, b1, ndmi2, bsi2, savi2, ndsi3, ndvi2, ndvi1, dvi3, bsi3, npcri1, evi1, evi3, ndsi2, ndwi1, savi1, si23, si33, ndsi1, b2, b3, dvi1, npcri2, ndmi1, bsi1, si42, savi3, bi2, b4, bi3, ndmi3, vssi3 npcri3, gdvi2, dvi2, int22, si22

RFE & RF Application

```
### RF Tuning

```{r}
# ensure the results are repeatable
set.seed(1787)
# load the library
library(mlbench)
library(caret)
# define the control using a random forest selection function
control <- rfeControl(functions=rfFuncs, method="cv", number=10)
# run the RFE algorithm
results <- rfe(train00[,1:73], train00[,74], sizes=c(1:73),
rfeControl=control)
# summarize the results
print(results)
# list the chosen features
predictors(results)
# plot the results
plot(results, type=c("g", "o"))
```
```

```
# RF00
```{r}
# Build the initial RF
library(randomForest)
set.seed(4543)
rf00 <- randomForest( v0 ~
bsi3+ndmi3+int12+si12+int13+si42+b3+gdvi2+npcri3+b8+dvi2+ndwi2+n
dwi3+gdvi3, data=train00, ntree=50, mtry =2,keep.forest=TRUE,
importance = TRUE, na.action = na.omit, keepForest = FALSE,
validation = "CV")
print(rf00)
plot(rf00)
imp <- importance(rf00)
varImpPlot(rf00, cex = 0.8)
```
```