HAL open science

Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches

Sanaz Zare, Ali Abtahi, Seyed Rashid Fallah Shamsi, Philippe Lagacherie

► To cite this version:

Sanaz Zare, Ali Abtahi, Seyed Rashid Fallah Shamsi, Philippe Lagacherie. Combining laboratory measurements and proximal soil sensing data in digital soil mapping approaches. CATENA, 2021, 207, pp.105702. 10.1016/j.catena.2021.105702 . hal-03338539

HAL Id: hal-03338539
https://hal.inrae.fr/hal-03338539

Submitted on 8 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title:

Combining laboratory measurements and proximal soil sensing data in Digital Soil mapping approaches

Authors:

Sanaz Zare [a,*], Ali Abtahi[a], Seyed Rashid Fallah Shamsi [b] and Philippe Lagacherie [c]

Affiliations:

[a] Department of Soil Science, College of Agriculture, Shiraz University, Shiraz, Iran

[b] Department of Natural Resources and Environmental Engineering, College of Agriculture, Shiraz University, Shiraz, Iran

[c] LISAH, Univ Montpellier, INRAE, IRD, Montpellier SupAgro, Montpellier, France

Corresponding author:

Sanaz Zare, sanaz_zare888@yahoo.com, Tel.: +989173203710, Postal Code: 71441-65186

Combining laboratory measurements and proximal soil sensing data in Digital Soil mapping approaches

Abstract

Digital soil mapping (DSM) products are limited in accuracy because of the lack of soil inputs. Soil sensing is a promising alternative to direct soil measurements that could provide much denser spatial samplings. Although using relevant and detailed soil sensing input in DSM is considered as vital to increase the prediction performances, there has been no studies in the literature that compare and develop the methods for integrating new sources of soil data that can be applied as inputs of DSM. This paper fills this gap on the example of mapping electrical conductivities from sites with laboratory measurements, in-field EM38MK2 measurements and spatially exhaustive covariates. Three different approaches are tested for putting in synergy real measurements and EM38MK2 measurements: i) EM38MK2 measurement considered as measured points, ii) EM38MK2 measurement used for building a new soil covariate and iii) EM38MK2 measurement considered as a soft data in a regression co-kriging approach. According to soil analysis's financial expenditure, choosing an optimal sample size to merge laboratory analysis and in-field EM38MK2 measurements as surrogate data was done on the best method. The results showed i) the utility of EM38MK2 data in DSM as a surrogate input data for mapping soil salinity ii) Regression co-kriging was the best method for integration and iii) The impact of EM38MK2 data on the gains of performance becomes greater and greater as the sizes of real measurements of soil salinity decrease. Hence, in other areas worldwide that soil sensing as alternative data is accessible, this research's future utilization could be possible as a promising way to tackle one of the essential constraints of DSM.

Keywords: Digital soil mapping, Soil sensing, Soil salinity, Remote sensing, Regression co-kriging, Quantile Random Forest

1. Introduction

Digital soil mapping (DSM) products are limited in accuracy because of the lack of soil inputs. Some recent trials showed that increasing the density of spatial soil sampling substantially increased the prediction performances of DSM models (Lagacherie et al., 2020; Somarathna et al., 2017; Wadoux et al., 2019a). However, direct observation of soil is costly which explains why most of the DSM activity is made from legacy data. Soil sensing is a promising alternative to direct soil measurements that, under some measurement conditions, could provide much denser spatial samplings. One can distinguish remote sensing (Mulder et al., 2011) and proximal soil sensing (Viscarra-Rossel et al., 2011).

The literature includes a lot of successful estimations of soil properties by various soil sensing technologies, such as VIS-NIR-SWIR spectroscopy (Ben-Dor et al., 2008; Gomez et al., 2008; Gomez and Coulouma, 2018; Lagacherie et al., 2008; Minasny et al., 2009; Viscarra Rossel et al., 2009); Gamma-ray spectroscopy (Buchanan et al., 2012; Spadoni and Voltaggio, 2013; Triantafilis et al., 2013; Zare et al., 2018); Ground penetrating radar (Abbaszadeh Afshar et al., 2016; Koyama et al., 2017; Lu et al., 2017; Tosti et al., 2013; Weihermuller et al., 2007); airborne hyperspectral imagery (Gholizadeh et al., 2018; Gomez et al., 2015; Hong et al., 2020; Nouri et al., 2017), and time-domain reflectometer (Arsoy et al., 2013; Bittelli et al., 2008).

EMI has become very popular among the available soil sensors for mapping the soil properties that affect ECa (Corwin and Scudiero., 2016), such as clay content, cation exchange capacity, water content, and pH (Corwin et al., 2003; Triantafilis and Lesch, 2005; Zare et al., 2015; Zhao et al., 2020). It is a valuable asset for mapping soil salinity, as an essential element of environmental surveillance and monitoring, using reliable approaches such as linear regression equation between ECa and ECe (Herrero and Hudnall., 2014; Amezketa and de Lersundi, 2008) and linear regression between ECe and calculated true electrical conductivity by inversion algorithm (Zare et al., 2015). Numerous literature that presented the high correlation between ECa and soil salinity (Corwin et al., 2003; Ding and Yu, 2014; Huang et al., 2014; Taghizadeh-Mehrjardi et al. 2016; Yao et al., 2012); inspired us that DSM can integrate EM38 prediction results of soil salinity to provide a more precise map.

It should be noted that most of these works intend to use soil sensing as a unique source of data, without considering any prior knowledge on soil distribution. Alternatively, soil sensing can also be considered as surrogate data for improving a soil mapping that is made from soil observations. Different methods could be used. The first one considers soil sensing as a covariate (Lagacherie and Gomez., 2018; Li et al., 2018; Taghizadeh-Mehrjardi et al., 2014; Zhang et al., 2020). As soil sensing measurements are usually not possible everywhere, soil sensing data should be pre-processed for getting a spatially exhaustive soil sensing covariate as required by the DSM approach. In this case, the empirical best linear unbiased prediction method (Zhang et al., 2020), kriging method (Taghizadeh-Mehrjardi et al., 2016), and regression kriging approach (Taghizadeh-Mehrjardi et al., 2014) performed to provide proximal soil sensing maps. Nevertheless, in sparse spatial sampling conditions, the regression kriging performances remain severely restricted (Vaysse and Lagacherie, 2015). Recently, Wang et al., (2021) created apparent soil electrical conductivity maps (ECa) using Random forests (RF) algorithms through environmental variables and electrical magnetic induction data. Another approach is to consider soil sensing as a soil site measurement, while considering that its uncertainty is greater than laboratory measurements. Such an approach was experienced when merging hyperspectral data with classical soil measurements (Walker et al, 2016). Co-kriging is considered as a possible method for doing that. It shows an improvement of the results. Therefore, the selection of an appropriate method for integrating new sources of soil data that can be applied as inputs of DSM models is crucial

Although using relevant and detailed soil sensing input in DSM is considered as vital to increase the prediction performances, there has been no studies in the literature that compare and develop the above-evoked alternatives of using soil sensing data in DSM approaches. This paper fills this gap on the example of mapping electrical conductivities from sites with laboratory measurements, in-field EM38MK2 measurements and spatially exhaustive covariates. Three different approaches are tested for putting in synergy real measurements and EM38MK2 measurements: i) EM38MK2 measurement considered as measured points, ii) EM38MK2 measurement used for building a new soil covariate, and iii) EM38MK2 measurement considered as soft data in a regression co-kriging approach. According to soil analysis's financial expenditure, choosing an optimal sample size to merge laboratory analysis and in-field EM38MK2 measurements as surrogate data was done on the best above-mentioned method. This part attempted to show the possibility of reducing ECe laboratory measurements in situations where EM38MK2 data exist.

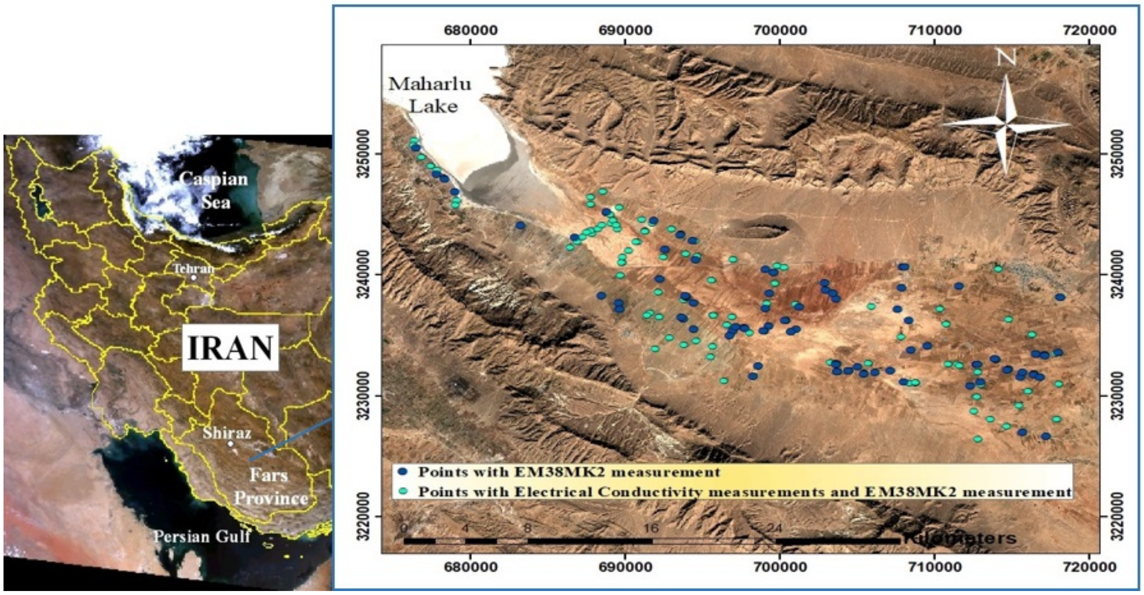

Fig 1: Study area in Southern Iran

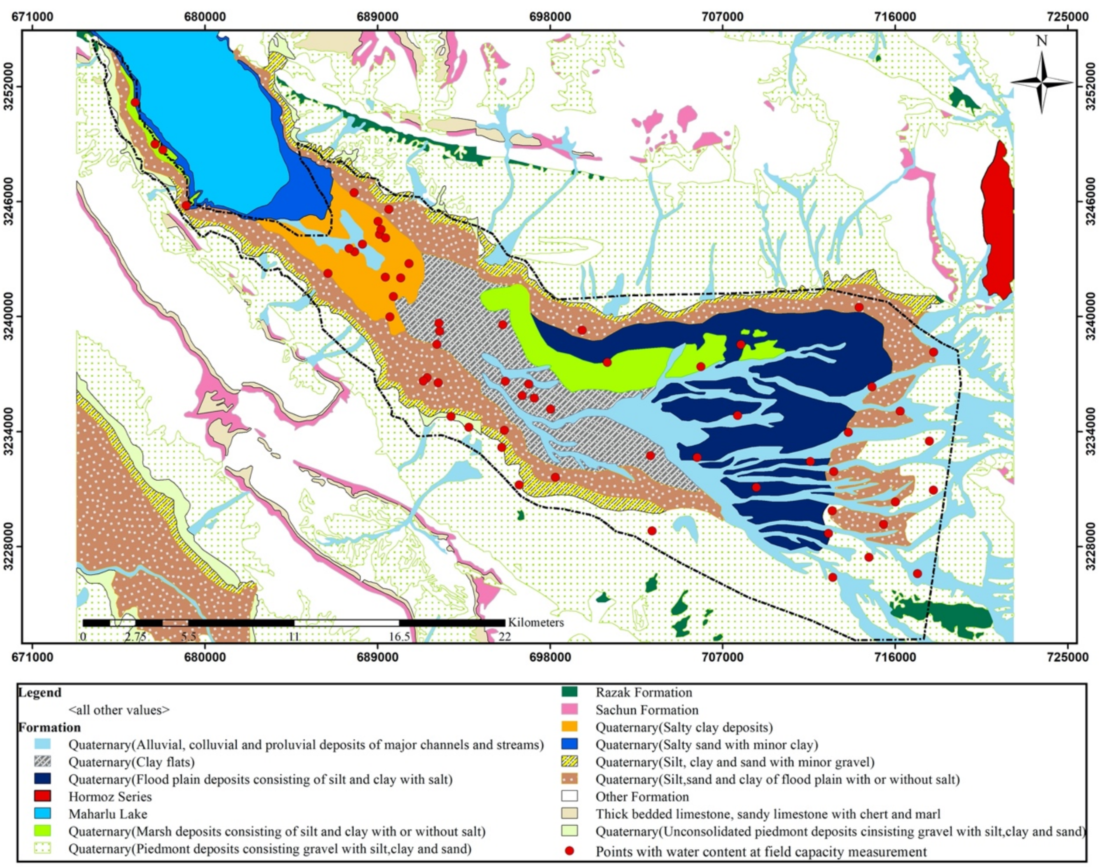

Fig. 2: Geological map of the region

2. Materials and Methods

2.1. Study Sites

The present study was carried out in Sarvestan region, near to saline-alkaline Maharlu Lake, is located in the southeast of Shiraz, Fras province, Iran (Fig.1), which is dominated by farmland and rangelands land cover and the soil's parent materials are highly calcareous (Abtahi, 1980; Khormali et al., 2003). The exposed geological formations are notably composed of Razak evaporites, Pabdeh-Gurpi shales and marls, Asmari-Jahrum limestone and dolomite, Sachun gypsiferous marls, Aghajari sandstone, Sarvak limestone, and Bakhtiari conglomerates (Fig.2). Two prominent salt domes (Hormuz formation) containing halite with a small amount of gypsum and other evaporite minerals are located in the southeast and northeast of the plain (Raeisi et al., 1996).

The study area's mean annual temperature and precipitation are 18 ºC, and 328.6 mm, respectively. This region contains plains with slight to severe salinity due to semi-lacustrine and lacustrine conditions and the outer margin with no salinity challenges (Abtahi, 1980). Factors such as high temperatures and salinity and alkalinity of shallow groundwater have been caused the formation of saline soils in the study region (Fallah Shamsi et al., 2013). The more intensive conditions throughout recent years include droughts, an increase in demand for water resources as well as the excessive use of chemical fertilizers (Zare et al., 2019). This region's main land uses include irrigated farming, dryland farming, rangeland, barren land, wetland, and urban. Pistachio, ficus, almond and olive trees, wheat, barley and maize are the dominant crops in the study area. The predominant plant community in the very saline and moderately salt-affected soils in this region are Salicornia Sp., Salsola Sp., Suaed Sp., Prosopis stephaniana, Alhagi camelorum, and some Gramineae (Abtahi, 1980). To manage and remediate, the traditional cropping patterns are changing to more adaptable ones that substitute water consuming crops and trees with salt-tolerant plant species (such as Barley and Pistachio). This recalls the necessity of monitoring ongoing salinization and developing procedures to identify and plan initial stages of soil salinity to provide information to ameliorate salt-affected soils by cost-effective and proper decisions (Metternicht and Zinck., 2008).

2.2. Data

2.2.1. Soil data sampling

2.2.1.1. Soil sampling and laboratory measurements

Soil sampling (Fig.1) was done in March 2019, in the time window of the EM38-MK2 survey, and when in the study area, the soil profile contains near to field capacity water content. The sampling approach followed the conditioned Latin hypercube method (Minasny and McBratney, 2006) employing spatial soil covariates that cover the most variation within the area for gathering 372 soil samples in 124 soil pits from the three equal soil depth ranges (0–0.3 m (topsoil), 0.3–0.6 m (subsurface) and 0.6–0.9 m (subsoil)) using a rotating auger. Employing a Global Positioning System (GPS) handset, the coordinates of soil samples were recorded. After delivering samples to the laboratory, the samples air-dried, sieved (2 mm) and the electrical conductivity of a saturated soil paste extract (ECe, dSm^{-1}) were determined (US Salinity Laboratory Method,1954). Moreover, soil moisture was determined gravimetrically, and in the topsoil and subsurface of 62 above-mentioned soil pits, water contents at field capacity (-33 kPa) were measured using a pressure plate (Dane and Hopmans, 2002) (Fig.2). Subsequently, the ratio of soil moisture to the water content at field capacity was calculated.

2.2.1.2. Apparent electrical conductivity data

EM38-MK2 device was employed to measure the apparent electrical conductivity (ECa, mSm−1). The EM38-MK2 implements simultaneous measurements of ground conductivity (Quad-Phase) and magnetic susceptibility (In-Phase) in vertical and horizontal dipole mode by couple transmitter-receiver coil detachment at 1 m and 0.5 m (Geonics Limited, 2009).

ECa was measured between 9th to 15th March 2019 when in the study area, the soil profile contains near to field capacity water content, which ensured reliable EMI signal data because EM38 survey in arid conditions is especially problematic as the conductance through the liquid pathway reduced (Corwin and Lesch.,2013; Corwin and Scudiero.,2016).

Using the conditioned Latin hypercube method (Minasny and McBratney, 2006) 214 points were selected for the EM38-MK2 survey (Fig.1) and the EM38-MK2 measurements were made at the location of the 124 sites with real measurements and the remaining (90) on new sites. The apparent electrical conductivity was measured in the vertical (v) and in the horizontal (h) mode that confirmed the recognition of variations in ECa to effective depths of 0.38 (MK2-h-0.5), 0.75 (MK2-h-1.0), 0.75 (MK2-v-0.5), and 1.5 m (MK2-v-1.0). The EM38-MK2 was nulled and calibrated according to the user manual (Geonics Limited, 2009) before each day and during the survey.

2.2.2. Spatial soil covariates

2.2.2.1. Digital elevation model (DEM) and derivatives:

A 10-m spatial resolution digital elevation model was provided from the National Cartographic Center of Iran (2014). From DEM, different terrain attributes including elevation, curvature, slope gradient, aspect, Multi-Resolution Valley Bottom Flatness (MRVBF), and Flow direction were acquired using SAGA GIS software (Conrad et al., 2015).

2.2.2.2. Maps

1:100,000 scale geological map to picture the spatial pattern of the parent material soil forming factors. This map was obtained from Geological Survey and Mineral Exploration of Iran.

A land use map has been produced by the Natural Resources Office of Fars province, updated using intensive field surveys and Google Earth image interpretation.

2.2.2.3. Remotely-sensed Data

Sentinel-1A, Sentinel 2A, Landsat-8 OLI/ TIRS satellite images, which are free of charge for users and proper for digital soil mapping investigations, were used in this research.

Sentinel-1A: The Sentinel-1A C-band SAR imagery with the advantage of being insensitive to water vapor or cloud cover, and a 12- day revisit cycle was acquired in Interferometric Wide swath mode (IW) with dual polarization, resulting in a VV and VH band for the image. The penetration capability of C-band radars is limited but slightly better than X-band. The Sentinel-1 toolbox in the SNAP 7.0 software was employed for the preprocessing, including radiometric calibration, thermal noise removal, and terrain correction with Shuttle Radar Topography Mission (SRTM-30m).

Sentinel-2: The cloud-free Sentinel-2 Image satellite with 13 spectral bands and 10, 20, 60 m spatial resolution in the visible, NIR, and SWIR spectrums was acquired from ESA Sentinel Scientific Data Hub within the time window of fieldwork and soil sampling. The Sentinel-2 Level 1C image was atmospherically corrected and was processed to atmospherically corrected bottom of atmosphere reflectance (Level 2A) using the Sen2Cor algorithm.

Table 1: The soil remote sensing attributes.

$NDVI = (NIR - R) / (NIR + R)$	(1)	Normalized Difference Vegetation Index (Rouse et al., 1974)
$NDWI = (NIR - SWIR_1) / (NIR + SWIR_1)$	(2)	Normalized Difference Water Index (Cheng et al., 2008)
$NDSI = (R - NIR) / (R + NIR)$	(3)	Normalized Difference Salinity Index (Khan et al., 2001)
$SI_1 = \sqrt{G \times R}$	(4)	Salinity Index (Khan et al., 2001)
$SI_2 = \sqrt{G^2 + R^2 + NIR^2}$	(5)	Salinity Index(Douaoui et al, 2006)
$SI_3 = \sqrt{G^2 + R^2}$	(6)	Salinity Index (Douaoui et al, 2006)
$BI = \sqrt{R^2 + NIR^2}$	(7)	Brightness Index (Khan et al., 2001)
$SRWI = (B_2) / (B_5)$	(8)	Simple Ratio Water Index (Maffei et al, 2007)
$CI = (SWIR1) / (SWIR2)$	(9)	Clay index (Carranza and Hale (2002))
$VARI = G - R/G + R - B$	(10)	Visible Atmospherically Resistant Index (Stow et al., 2005)
$RVI = (NIR) / (R)$	(11)	Ratio Vegetation Index (Pearson and Miller 1972)
$MTVI = 1.2(1.2(800nm - 550nm) - 2.5(670nm - 550nm))$	(12)	Modified Triangular Vegetation Index (Karnieli et al., 2001)
$OSAVI = (NIR - RED) / (NIR + RED + 0.16)$	(13)	Optimized Soil Adjusted Vegetation Index (Rondeaux et al. (1996)
$SAVI = [(NIR - red) / (near\ infrared + red + L)] * (1 + L)$	(14)	Soil adjusted vegetation index (Huete .,1988)
$DVI = (NIR - RED)$	(15)	Difference vegetation index (Tucker., 1979)
$EVI = 2.5(NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1)$	(16)	Enhanced vegetation index (Huete .,2002)
Tasseled cap transformation	(17)	Huang et al. (2002)
GLCM mean	(18)	Haralick et al. (1973)
GLCM correlation	(19)	Haralick et al. (1973)
GLCM Variance	(20)	Haralick et al. (1973)
Median Filter	(21)	Haralick et al. (1987)

Landsat 8 : Landsat 8 with 16 days revisiting frequency, carries the Operational Land Imager (OLI) and thermal Infrared Sensor (TIRS) which collect data in VIS, NIR and SWIR bands with 30 m, the panchromatic band with 15 m, and TIR bands with 100 m spatial resolution.

Cloud free image of Landsat-8 OLI/ TIRS was obtained on the 11th of March, 2019 from the USGS Earth Explorer website (https://earthexplorer.usgs.gov). ENVI5.3 was employed for the atmospheric correction by FLAASH algorithms (Cooley et al., 2002) and radiometric calibration.

We acquired the tasseled cap transformation as a practical data dimensionality reduction approach(Crist and Cicone.,1984), several image textural features, soil, and vegetation transformations by utilizing Eqs.1 to 16 (Table 1). The textural variables were provided using the grey level co-occurrence matrix (GLCM) with the 5*5 kernel size (Haralick et al., 1973).

The spatial soil covariates explained above were registered to a common grid of 30 m cell size.

2.3. Methods

2.3.1. Mapping Model: Quantile Regression Forest

For the prediction of soil ECe and ECa, the Quantile regression forest (QRF) algorithm (Meinshausen., 2006) was applied. Breiman (2001) and Meinshausen (2006) reported the comprehensive explanation of random forests and quantile random forests, respectively.

QRF is a non-parametric and robust ensemble learning method that has been increasingly applied to DSM (Dharumarajan et al., 2020; Liu et al., 2020; Szatmári and Pásztor., 2019; Vaysse and Lagacherie, 2017). Similar to the random forest (RF), the QRF algorithm comprises numerous tree predictors with randomly split nodes. RF uses bagging (bootstrap aggregating) to improve the stability of results and decrease the risk of overfitting. RF Predictions are usually constructed from the mean of predicted values created from numerous decision trees. In contrast, QRF considers the response variable's spread of values at each node and infer estimates for conditional quantiles, prediction intervals, or other statistics from the distribution (Dobarco et al.,2019; Meinshausen, 2006; Vaysse and Lagacherie, 2017). If there are extreme values in the samples applying the sample mean in the leaf node may result in biasness (Gyamerah et al., 2020), therefore the median value was used for point prediction in the QRF model to enhance the accuracy of the prediction.

For the present study, the ranger package (Wright and Ziegler, 2017) as the fast implementation of RF especially fitted for high dimensional data and the tuneRanger package (Probst et al., 2018) were applied for operating the QRF models in R software.

Hyperparameters of Random Forest algorithms require to be tuned to gain bias-reduced assessment and better performances (Probst et al., 2018). tuneRanger package helps to identify the best RF hyperparameters for running the model using sequential model-based optimization (Hutter et al., 2011; Jones et al., 1998; Probst et al., 2018). Regarding computing costs, 100 repetitions showed to be suitable for a fine convergence to an optimized adjustment (Lagacherie et al., 2020).

2.3.2. Feature screening

Determining the most important covariates to obtain the most accurate predictions is the purpose in numerous machine learning researches. Random Forest is not affected by a vast number of covariates; also more covariates than measurements can be applied (Hengl et al.,2018) and with a more expansive selection, the probabilities of having the most suitable covariates accessible to the algorithm will be enhanced (Khaledian and Miller, 2020). In terms of prediction, Random Forest can handle the correlated covariates, using bootstrap and an out-of-bag (OOB) strategy. Nonetheless, the covariate importance grade would be influenced if the covariates that are highly correlated to the really influential covariates getting picked up together and over-selected (Huang and Boutros., 2016; Strobl et al.,2007). Consequently, the most important covariates were selected using the Pearson correlation coefficients, principal component analysis (PCA), and QRF. The PCA explores underlying properties that summarize a group of highly correlated properties. In this regard, the Pearson correlation coefficients and the PCA of covariates were determined, and some covariate with similar information were omitted.

Then, the QRF was trained on the filtered covariates, and variables importance were ranked and further the least important covariates were removed. Finally, QRF was built using 41 selected covariates among 500 initially defined. In this study, a permutation-based method (Breiman, 2001) was used to measure the factor's importance. In this method, the variable is recognized as important if it positively influences the prediction's performance (Probst et al.,2018).

2.3.3. Ordinary Co-Kriging

The co-kriging was applied as the best linear unbiased estimator, owning minimum estimating error variance (Wackernagel, 1995) which integrates a sparely measured primary variable with a more densely secondary variable to employs the cross-correlation of them (Grunwald, 2006). Ordinary co-kriging was applied using the package GSTAT R (Pebesma, 2004). The predicted soil properties employing CK can be formulated as Eq.17 (Li and Yeh., 1999).

$$\widehat{f(x)} = \sum_{i=1}^{n} \lambda_i f(x_i) + \sum_{j=1}^{m} \lambda_j' f(x_j) \qquad \text{(Eq. 17)}$$

Where $\widehat{f(x)}$ is the predicted value of soil properties, *n* and *m* are the numbers of locations with observed soil properties and secondary variables respectively, $f(x_i)$ and $f(x_j)$ are; respectively, observed values of the soil property at location i and of the secondary variable at location j and λ_i and λ_j' are the CK weights of those observed values. Where the $\lambda_{i'}$ s and λ'_j 's solve the consequent cokriging method with ni+nj+2 equations to confirm the minimization of the MSE and unbiasedness:

$$
\begin{cases}
\sum_{i'=1}^{n_i} \lambda_{i'}\ C_{i,i'}^{(ZZ)} + \sum_{j=1}^{n_j} \lambda'_j\, C_{i,j}^{(ZY)} - \sum_{K=1}^{P} \mu_K = C_{i,0}^{(ZZ)} & for\ i = 1, \dots, n_i \\
\sum_{j'=1}^{n_j} \lambda'_{j'}\ \ C_{j,j'}^{(YY)} + \sum_{i=1}^{n_i} \lambda_j\, C_{i,j}^{(ZY)} - \sum_{K=1}^{P} \mu_K = C_{j,0}^{(ZY)} & for\ j = 1, \dots, n_j \\
\sum_{i=1}^{n_i} \lambda_i \ = 1 \qquad and \qquad \sum_{j=1}^{n_j} \lambda'_j \ = 0 &
\end{cases} \qquad \text{(Eq. 18)}
$$

2.3.4. Regression Kriging

Regression kriging incorporates the spatial dependency in the regression residuals into the kriging procedure (Hengl et al., 2004). This method combines the relationships between soil properties and spatial soil covariates through different linear and non-linear regression models with kriging of the regression residuals (Hengl et al., 2007; Vaysse and Lagacherie, 2017). The regression kriging (RK) of soil properties for location x, $Z_{RK}(x)$, is defined as the sum of regression estimate $Z_r(x)$ and the estimate of spatially correlated residual values $\varepsilon_{ok}(x)$ applying the subsequent equation (Hernndez-Stefanoni et al., 2011) (Eq.19):

$$Z_{RK}(x) = Z_r(x) + \varepsilon_{ok}(x) \qquad \text{(Eq. 19)}$$

2.3.5. Using EM38MK2 data in DSM

In this study, the EM38MK2 data were used as surrogate soil inputs following three approaches that are detailed below.

2.3.5.1. Approach 1: " EM38MK2 as new measured sites"

As a first approach, we aim adding the sites measured with EM38MK2 to the set of laboratory measurements of ECe. A pedotransfer function that convert EM38MK2 values into real values of

ECe was first calibrated onto the 120 sites having the two measurements. Finally, a QRF model was built from all the sites data to find the relationships between ECe and the environmental covariates at each depth interval (Fig.3). In this case, the estimations of a pedotransfer function at the point with only EM38MK2 data were supposed to be accurate enough to be considered as real measurements of a soil ECe. This potentially substitutes a sparsely measured objective variable with a more dense soil ECe data, which has the benefit to improve covering the changes in soil ECe in the study area. Nevertheless, these extra approximations carry uncertainties that can influence the model's result.

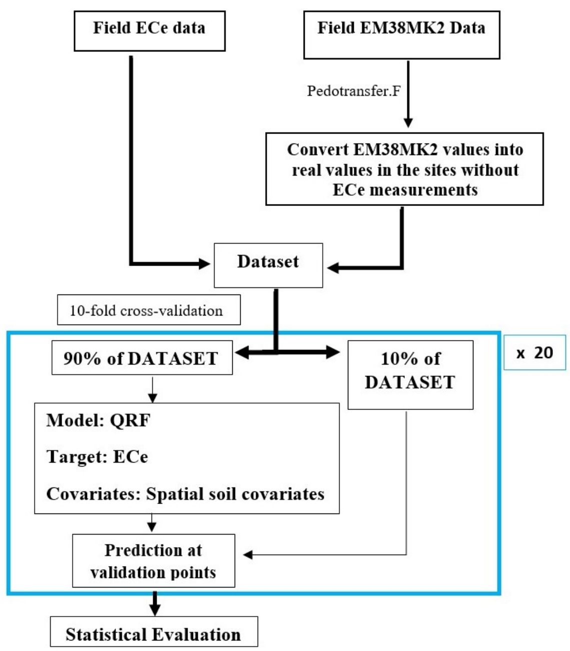

Fig. 3. Flowchart of the first approach

Stepwise multiple linear regression (SMLR) method was implemented to model correlation among soil ECe and the MK2-h-0.5, MK2-h-1.0, MK2-v-0.5, and MK2-v-1.0 data. The model was tested for the multi-co-linearity of the selected independent variables. Avoiding collinearity due to closely correlated variables, allowed us to achieve more precise models because applying the collinear variables decreases the model's accuracy. In this analysis, the most suitable models were chosen based on the criteria with higher R2, lower RMSE, and employing Variance Inflation Factor and tolerance values (Kutner et al., 2005).

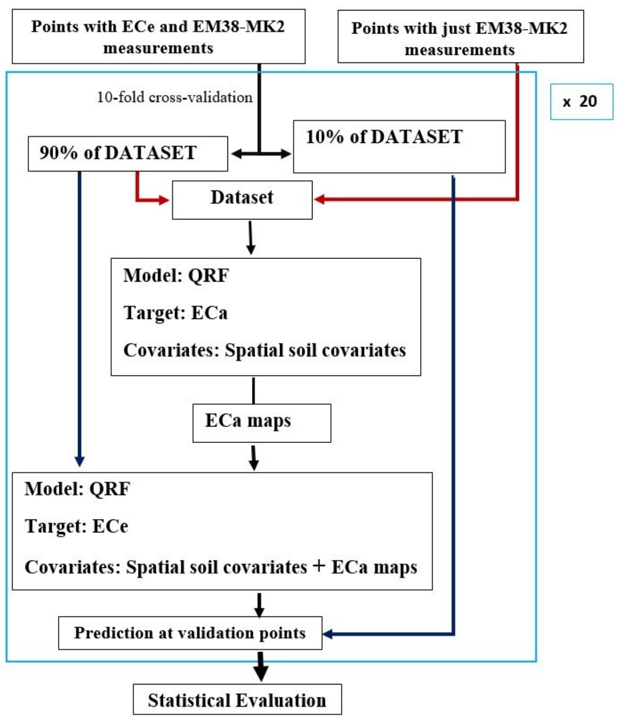

Fig. 4. Flowchart of the second approach

2.3.5.2. Approach 2 : “EM38MK2 as a new soil covariate”

In this second approach, the EM38MK2 data were used to produce exhaustive ECa maps that were added to the set of covariates applied by the DSM model that was built from the 120 sites with real measurements of EC. Four ECa maps were produced (MK2-h-0.5, MK2-h-1.0, MK2-v-0.5, and MK2-v-1.0) on the basis of EM38MK2 measurements and the environmental covariates by calibrating QRF from the 210 sites with EM measurements. The most important ECa map regarding Pearson correlation coefficients, collinearity between MK2-h-0.5, MK2-v-0.5, MK2-h-1.0, MK2-v-1.0 measurements, and QRF as feature selection algorithms was selected. This new covariate was added to the set of covariates and a DSM model was calibrated using the 120 sites with real measurements (Fig.4).

2.3.5.3. Approach 3 : "EM38MK2 as soft data of EC"

The third approach followed a regression co-kriging approach. A QRF model using as calibration data the 120 sites with real measurements was first built. Their residuals were calculated both on the sites with real measurements of EC (hard data) and on the sites with EC estimates obtained as described in section 2.3.5.1. (soft data). Regression co-kriging approach of the residuals using the former as hard data and the latter as soft data was performed. The final predictions were calculated by adding the cokriged residuals to the ECe values predicted by the QRF model. In this approach, we dealt with the uncertainties related with ECe estimations from ECa evoked for the first approach by considering residuals in regression co-kriging (Fig.5).

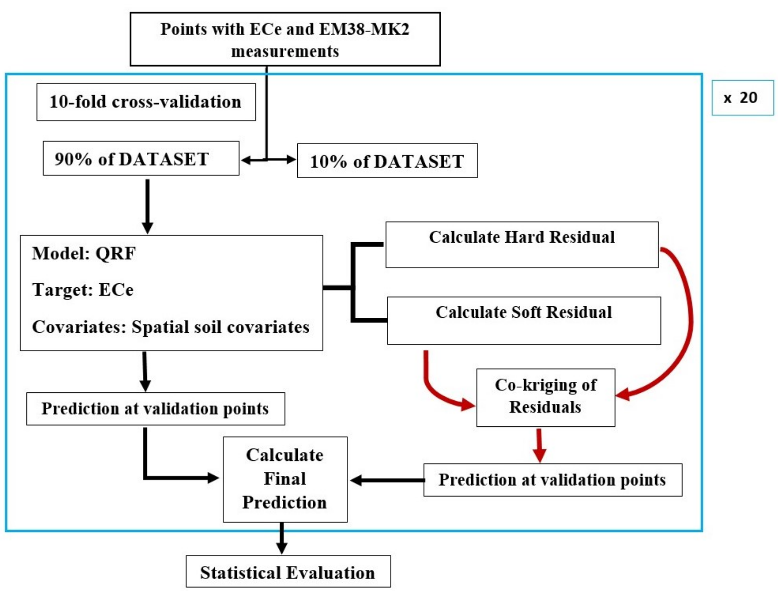

Fig. 5. Flowchart of the third approach

The three above-described approaches were compared to a baseline approach (approach 0) that consists in simply calibrating QRF from the 120 sites with ECe measurements, without considering the EM measurements.

2.3.6. Comparing soil inputs with different sizes of ECe laboratory measurements

In the perspective of providing accurate soil maps with a fair expense and time, we explored the possibilities of reducing the costly ECe laboratory measurements in situations where may exist spatial sampling of EM38MK2 measurements. This part of the study was done only for the best out of the three above-mentioned approaches.

In this respect, we produced two new spatial sampling of ECe measurements by sampling 50% and 75% of the sites with real measurements (120 sites) using a stratified random sampling method. EM38MK2 measurements were substituted to ECe at non-selected locations. Consequently, the first dataset contained 60 sites with real measurements and 120 sites with in-field EM38MK2 measurements and the second dataset contained 90 sites with real measurements and 150 sites with in-field EM38MK2 measurement.

2.3.7. Evaluation Protocol

All the three tested approaches were evaluated from the 120 sites with real measurements.

In order to use all the data and increase the robustness of the evaluation, the total dataset was divided randomly into ten folds with the same size on the basis of the k-fold cross-validation (k = 10) method with 20 times replication. This strategy involved employing the first fold as an evaluation set and fitting the model on the left k-1 fold and k times was iterated until all folds had been utilised as the evaluation set. In this way, all three horizons' predictions of all soil data were compared with the observed data for the entire dataset. It should be noted that the production of

EC maps from EM measurements performed in approach 2 was included in the cross-validation loop. This ensured that the ECa maps used as covariates were not produced using EM measurements performed at the same locations as the validation sites, which guarantee an independent (and unbiased) evaluation .

The model's performances were evaluated, using mean square error skill score (SSmse) (Nussbaum et al., 2018), root mean squared error (RMSE), normalized root mean square error (nRMSE), where RMSE is normalized by dividing by the means of the observed data, and mean error (ME). SSmse has the same interpretation as the R^2 and is the percentage of variance that explained by the model. ME and RMSE also displayed estimation errors; nevertheless, RMSE has more sensitivity to outliers (Taylor,1997). Furthermore, we calculated the ratio of the performance to interquartile distance (RPIQ = (Q3 − Q1)/RMSE), where Q1 and Q3 are the first and third quartiles (Khaledian and Miller, 2020), considering the reliability of the prediction: very poor model (RPIQ < 1.4), fair (1.4 ≤ RPIQ < 1.7), good model (1.7 ≤ RPIQ < 2.0), very good models (2.0 ≤ RPIQ ≤ 2.5), and excellent models (RPIQ > 2.5).

For the models built from reduced sets of measurements, the number of samples in the calibration set was 60 and 90 sites, and the evaluation was conducted over the same sample size set (120 measurement sites). Consequently, in the K-fold cross-validation, for the former calibration set, the first fold plus 60 sites and for the latter calibration set, the first fold plus 30 sites were used as a validation set, and the model was fit on the rest K-1 folds. The MSE was measured on the sites in the held-out fold and 60 and 30 sites respectively and then RMSE, SSmse, and ME were calculated.

3. Results

3.1. Statistical analysis

3.1.1. Exploratory Data Analysis

Table 2 indicates the descriptive statistics of the soil ECe at the different depths before and after normalization and the ratio of soil moisture to the water content at field capacity. The 0.05 level of significance (Kolmogorov–Smirnov) was used for assessing the normality of distributions. Soil ECe showed positively skewed distributions and so were subjected to log-transformation, which agrees with the most frequently reported results. Hence, the model was created from the Log-transformed data, and later the predicted EC was achieved by back transformation of data. With respect to the mean and Q1 to Q3 of ECe, the salinity varied from non-saline (< 2 dS m−1) to extremely saline (> 16 dS m−1) at all depths. In general, salinity was higher in the subsurface and subsoil.

To present soil ECe changes in vertical direction, the topsoil samples were classified into the common soil salinity classes (Richards, 1954): 0-2 dS m−1 (S1), 2-4 dS m−1 (S2), 4-8 dS m−1 (S3), 8-16 dS m−1 (S4), and >16 dS m−1 (S5). Based on the soil salinity classes of the topsoil samples, the box plots were calculated for all layers. These plots showed that the mean values of ECe in subsurface and subsoil were higher than topsoil samples in all categories. By increasing in depth, approximately, soil categories seem to shift toward more saline ones, i.e., S1 in topsoil to S2 the subsurface, even S3 in the subsoil (Fig.6).

The ratio of soil moisture to the water content at field capacity showed that the water content was near field capacity with the mean value of 0.75 and 0.76 for topsoil and subsurface layers, respectively (Table 2), and electrical conductance was not limited by inadequate soil moisture in the study area. These results suggest that rising shallow saline and alkaline groundwater, high temperature, and slightly salt leaching from the topsoil in winter result in salt accumulation in the subsurface because of insufficient drainage.

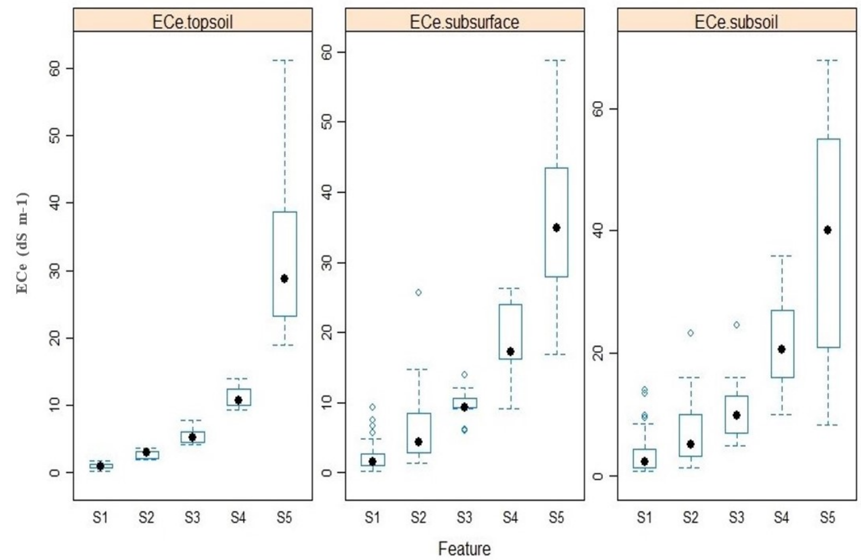

Fig. 6. Calculated ECe (dS m−1) box plot based on soil salinity classes of topsoil samples.

Table 2: Descriptive statistics of soil ECe (dS m-1) and water content (represented as percent of field capacity) in the study area

Layer (cm)	Min	Max	Mean	SD	CV%	Q1	Median	Q3	Skewness	Kurtosis	Kolmogorov -Smirnov
ECe 0-30	0.16	61.22	8.48	13.02	1.53	0.95	2.1	10.07	2.22	4.71	0.00
ECe 30-60	0.10	58.90	11.39	13.94	1.22	1.52	4.58	16.99	1.62	1.93	0.00
ECe 60-90	0.60	68.00	12.37	15.07	1.21	2.13	5.5	16	1.88	3.08	0.00
Log ECe 0-30	-0.78	1.78	0.48	0.63	1.31	-0.02	0.32	1.00	0.36	-0.89	0.08
Log ECe 30-60	-1.00	1.77	0.70	0.60	0.86	0.18	0.66	1.23	-0.07	-0.83	0.48
Log ECe 60-90	-0.22	1.83	0.78	0.54	0.69	0.32	0.73	1.20	0.04	-0.96	0.77
water content (% of FC) 0-30	0.66	0.84	0.75	0.04	0.06	0.74	0.76	0.79	-	-	-
water content (% of FC)30-60	0.65	0.87	0.76	0.06	0.08	0.70	0.75	0.83	-	-	-

FC: Field capacity

3.1.2. Relationship between ECe and ECa data

For all the ECe and ECa data, Pearson correlation coefficients were calculated, as can be seen in Table 3, and indicated that the vertical (v) mode of EM38MK2 is strongly correlated with the horizontal (h) mode. According to the results using EM38MK2 data will be informative to predict ECe and closely reflects the spatial distribution ECe (Corwin and Lesch., 2005). This could be relevant to the fact that the salt content mainly affected ECa in the saline area's soil (Rhoades, 1990). The MK2-v-1.0 readings were used to produce exhaustive ECa map in the second approach with regard to the most significant correlation value between Log ECe and MK2-v-1.0 at all depths and collinearity between MK2-h-0.5, MK2-v-0.5, MK2-h-1.0,and MK2-v-1.0 measurements. In addition, feature screening using QRF confirmed that the MK2-v-1.0 was the most important covariate.

In order to evaluate the effect of soil moisture on ECa, the correlation coefficients between ECa readings and soil moisture were calculated (Table 3). The most significant correlation (r-value) was obtained between MK2-v-1.0 and soil moisture in the subsoil layer (0.44), followed by the subsurface (0.40) and topsoil layer (0.37). These results revealed that salinity is the soil feature that controls the ECa measurement and could be applied to predict ECe at all depths in the study area with regard to the larger correlation value between ECe and ECa (Zhao et al., 2020). In addition, the correlation trend of soil moisture with the soil depth may be relevant to the average values of soil moisture in the subsoil layer (23.55%), which was more than the subsurface (20.53%) and the topsoil layer (19.63%).

Table 3: Pearson coefficients (r) between the ECe (dS m-1), ECa data and Soil Moisture.

Layer (cm)	ECe 0-30	ECe 30-60	ECe 60-90	MK2-h-0.5	MK2-v-0.5	MK2-h-1.0	MK2-v-1.0	Log ECe 0-30	Log ECe 30-60	Log ECe 60-90	SM 0-30
ECe 0-30	1	0.91**	0.77**	0.87**	0.87**	0.87**	0.86**	0.84**	0.73**	0.66**	0.27**
ECe 30-60	0.91**	1	0.86**	0.87**	0.88**	0.88**	0.88**	0.86**	0.85**	0.78**	0.30**
ECe 60-90	0.77**	0.86**	1	0.91**	0.93**	0.93**	0.92**	0.76**	0.75**	0.85**	0.33**
MK2-h-0.5	0.87**	0.87**	0.91**	1	0.98**	0.98**	0.95**	0.75**	0.70**	0.72**	0.32**
MK2-v-0.5	0.87**	0.88**	0.93**	0.98**	1	0.99**	0.98**	0.77**	0.72**	0.75**	035**
MK2-h-1.0	0.87**	0.88**	0.93**	0.98**	0.99**	1	0.98**	0.77**	0.72**	0.75**	0.35**
MK2-v-1.0	0.86**	0.88**	0.92**	0.95**	0.98**	0.98**	1	0.79**	0.75**	0.78**	0.37**
Log ECe 0-30	0.84**	0.86**	0.76**	0.75**	0.77**	0.77**	0.79**	1	0.92**	0.80**	0.30**
Log ECe30-60	0.73**	0.85**	0.75**	0.70**	0.72**	0.72**	0.75**	0.92**	1	0.86**	0.31**
Log ECe60-90	0.66**	0.78**	0.85**	0.72**	0.75**	0.75**	0.78**	0.80**	0.86**	1	0.37**
SM 0-30	0.27**	0.30**	0.33**	0.32**	0.35**	0.35**	0.37**	0.30**	0.31**	0.37**	1
SM 30-60	0.34**	0.40**	0.40**	0.39**	0.40**	0.40**	0.40**	0.35**	0.36**	0.39**	0.66**
SM 60-90	0.36**	0.47**	0.47**	0.41**	0.42**	0.43**	0.44**	0.38**	0.42**	0.47**	0.60**

SM:Soil Moisture, **and * significant at the 0.01 and 0.05 level (2-tailed) respectively.

To establish calibration between ECa and ECe from soil samples, we applied SMLR between ECa data as independent variables and ECe as dependent variables for each depth of investigation. The SLMR was performed to the data from 120, 90 and 60 sampling sites and the results are summarized in Table 4. Fig. 7 showed the relationship between ECe (120 sites) and ECa of soil from the SMLR model. All the written regression equations meet the basic assumptions, including no or little multicollinearity among explanatory variables. According to the results, the recorded data in the 1.0-m vertical orientation allowed more reliable models (Heil and Schmidhalter., 2015) at the different depths and the models were satisfactory as regards calibration and thus the prediction of ECe. Besides, the correlation values between ECe and ECa were ranked as MK2-v-1.0, MK2-h-1.0 (MK2-v-0.5), MK2-h-0.5 in all the soil layers, which can be related to an increase in average soil moisture with depth.

Table 4: Summary of SMLR relationships between measured ECe and ECa (n = 120, 90, 60)

		LogECe = a + b (MK2-v-1.0)		
Layer (cm)	n	a	b	R^2
ECe 0-30 cm (dS m-1)	120	0.092	0.003	0.63
ECe 30-60 cm (dS m-1)	120	0.349	0.003	0.57
ECe 60-90 cm (dS m-1)	120	0.457	0.002	0.61
ECe 0-30 cm (dS m-1)	90	0.096	0.003	0.63
ECe 30-60 cm (dS m-1)	90	0.334	0.002	0.56
ECe 60-90 cm (dS m-1)	90	0.460	0.002	0.63
ECe 0-30 cm (dS m-1)	60	0.01	0.072	0.60
ECe 30-60 cm (dS m-1)	60	0.24	0.079	0.56
ECe 60-90 cm (dS m-1)	60	0.41	0.060	0.56

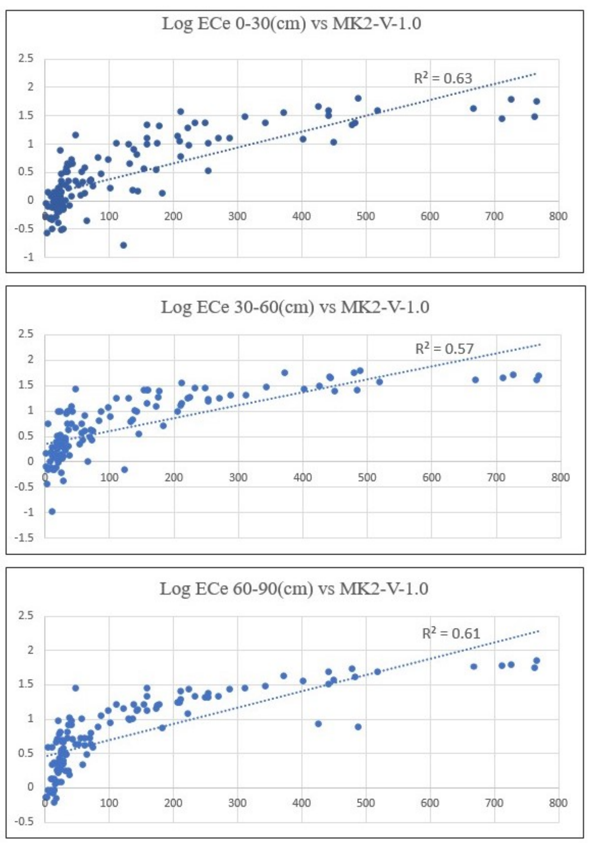

Fig. 7. Plot of coefficient of determination (R2) achieved between the Log ECe (dS m−1) and MK2-V-1.0

Table 5: Performances of the different approaches

		R^2	RMSE	nRMSE	ME	RPIQ
Approache 1	ECe 0-30 cm (dS m-1)	0.67	7.55	0.89	-2.38	1.21
Approach 2	ECe 0-30 cm (dS m-1)	0.72	7.61	0.89	-2.66	1.20
Approach 3	ECe 0-30 cm (dS m-1)	0.76	6.26	0.73	-0.32	1.46
Base Approach	ECe 0-30 cm (dS m-1)	0.69	7.21	0.85	-1.63	1.26
Approache 1	ECe 30-60 cm (dS m-1)	0.68	7.95	0.69	-2.09	1.95
Approach 2	ECe 30-60 cm (dS m-1)	0.71	7.76	0.68	-2.22	1.99
Approach 3	ECe 30-60 cm (dS m-1)	0.79	6.61	0.58	-0.13	2.34
Base Approach	ECe 30-60 cm (dS m-1)	0.70	7.52	0.66	-1.57	2.06
Approache 1	ECe 60-90 cm (dS m-1)	0.71	8.47	0.68	-2.50	1.64
Approach 2	ECe 60-90 cm (dS m-1)	0.75	8.06	0.65	-2.08	1.72
Approach 3	ECe 60-90 cm (dS m-1)	0.77	7.72	0.62	-0.08	1.80
Base Approach	ECe 60-90 cm (dS m-1)	0.73	7.94	0.64	-1.35	1.75

3.2. Prediction of spatial distribution of ECe

Table 5 indicates the performances for the three tested approaches for mapping electrical conductivities, through the cross validation procedure for different depth intervals. The first approach which relied on using EM38MK2 as measured points, showed the lowest performance and did not bring any improvement of the baseline approach. The second approach which use a spatially exhaustive ECa map, presented only a very slight improvement from the baseline approach. Conversely, the third approach which use Regression cokriging, improved significantly the performances compared to the baseline approach, especially for the subsurface soil layers (30-60 cm). In terms of RMSE, the approaches resulted in approximately the same prediction accuracy; which is related to the dependency of RMSE to the observed data's range. The RMSE values for the topsoil, subsurface, and subsoil layers were, respectively, 6.26, 6.61, 7.72 dS m−1, which are acceptable regarding the wide range of ECe (61.06, 58.8, 67.4 dS m−1) in the study area. The

RPIQ values ranged between 1.46 to 2.34, which exhibited that the third approach were accurate concerning the equivalent ranges of dataset spread. In addition, the predicted ECe by the third approach was, in general, unbiased given the small ME. The third approach showed an increasing trend in R2 and RPIQ with increasing depth to subsurface layer, and a reverse trend for nRMSE and ME. Furthermore, the subsurface layer's prediction performances outperformed the subsoil layer concerning the R2, nRMSE, RMSE and RPIQ.

Table 6: Performances of approach 3 with the different sample size

		Approach 3			Base Approach		
ECe (dS m-1)	n	R^2	RMSE	ME	R^2	RMSE	ME
ECe 0-30 cm	n:120	0.76	6.26	-0.32	0.69	7.21	-1.63
ECe 30-60 cm		0.79	6.61	-0.13	0.70	7.52	-1.57
ECe 60-90 cm		0.77	7.72	-0.08	0.73	7.94	-1.35
ECe 0-30 cm	n:90	0.64	9.28	-1.40	0.56	9.60	-1.8
ECe 30-60 cm		0.70	8.76	-1.18	0.62	9.15	-1.47
ECe 60-90 cm		0.74	8.58	-1.03	0.69	8.94	-1.19
ECe 0-30 cm	n:60	0.48	10.73	-1.52	0.39	11.38	-1.69
ECe 30-60 cm		0.54	10.78	-3.00	0.44	11.29	-3.04
ECe 60-90 cm		0.49	11.9	-4.25	0.40	12.88	-4.48

3.3. Effect of calibration models with different sample set sizes

According to the financial expenditure of soil analysis, choosing an optimal sample size to merge laboratory analysis and in-field EM38MK2 measurement as a surrogate data, was done on the third approach as the best above-mentioned method. Table 6 summarizes R2, RMSE and ME values, resulting from the approach validations for soil depths regarding the approach's type and sample's size. Table 6 illustrate that prediction accuracy improves with the increasing sampling size for all approaches and soil depths.

Comparison of the large size (120 soil sample) dataset models' accuracy with a medium size (90 soil sample) and small size dataset (60 samples) models' accuracy, showed that with decreasing sample sizes, differences between the third approach and base approach predictions increased. Nevertheless, the decreasing rate in the model's accuracy differs, and the highest reduction happened in the subsurface layer (0.3–0.6 m). This results revealed the importance of merging EM38MK2 data in the sparse dataset to cover the variation of the target variables in the study region, especially when there is a lack of intensive field data.

3.4. Spatial distribution of soil salinity

Spatial distributions of soil salinity content as mapped by approach three are shown in Fig.8. The main spatial distribution patterns of soil salinity in all soil depths revealed the directional reduction of the soil salinity from the coastal lake area to the further away region, and therefore at the outside margin, soil salinity did not affect the normal plant's growth. This distribution may be attributed to the different environmental and human-induced elements, notably comprising groundwater level, topography, drainage, saline irrigation water, and soil management manners.

4. Discussion

4.1. Added value of EMI data

The SMLR equations for converting EM38MK2 data into ECe data, in order to predict ECe from EMI data in the sites without soil sampling, conveyed uncertainties that could be explained by the fact that ECa readings were affected by diverse soil properties, for instance, soil texture and soil moisture, although salinity is commonly the soil feature that controls the ECa measurement

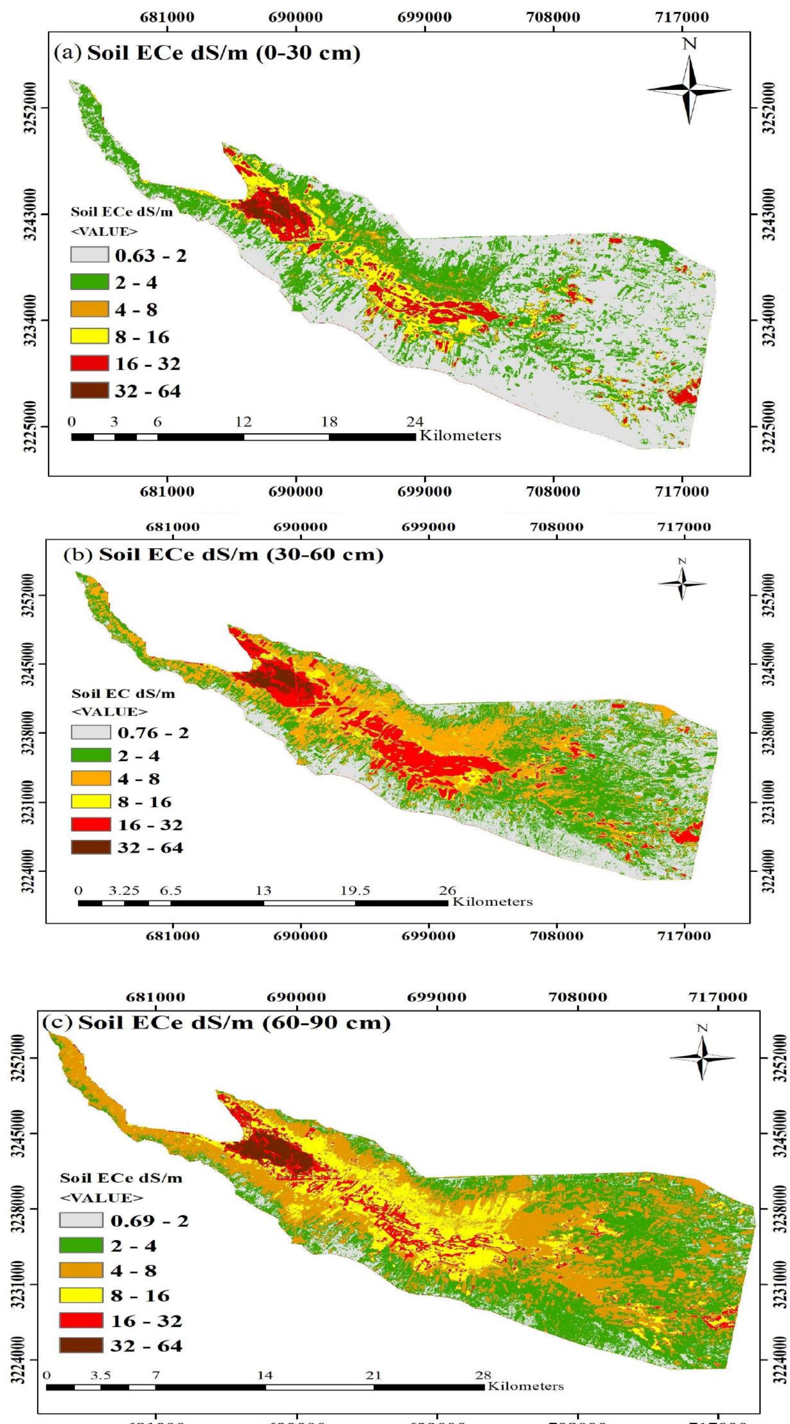

Fig.8. Predicted maps of ECe (dS m−1) for the entire study region: (a) 0–30 cm, (b) 30–60 cm, (c) 60–90 cm.

(Lesch et al.,1998; Rhoades et al.,1990; Slavich 1990; Taghizadeh-Mehrjardi et al.,2014). In addition, this may be related to the quite different volumes of soil measured by the EM38MK2 survey and the soil samples provided using a rotating auger from the three equal soil depth ranges (0–0.3 m, 0.3–0.6 m, and 0.6–0.9 m) to measure ECe (Martini et al., 2013; Rhoades et al.,1990). Calibration of ECa measurements using linear regression model were already used by Taghizadeh-Mehrjardi et al. (2014; 2016), Ding and Yu (2014) and numerous studies documented these approaches (Feikema and Baker., 2011; Herrero and Hudnall., 2014; Rhoades et al., 1990; Slavich., 1990; Triantafilis et al., 2000; Triantafilis and Buchanan., 2010; Yao and Yang., 2010). Other researchers (Khongnawang et al., 2019; Zhao et al., 2020) have reported that where a direct linear regression model between soil properties and ECa can not be set up, ECe may be mapped by creating a linear regression among estimates of true electrical conductivity with soil properties. According to the results, the ECa map provided by the readings of the 1.0-m vertical configuration is the most important covariates in approach 2, which can be related to an increase in average soil moisture with depth. Heil and Schmidhalter (2015), Taghizadeh-Mehrjardi et al.,(2014) and Wang et al., (2021) reported the similar results for soil texture and ECe predictions at surface and subsurface of the soil.

The results indicated that inclusion of soil ECa had more advantages for enhancing the ECe prediction in the subsurface layer (0.3–0.6 m) which can be related to the effective depth of the instrument. Heil and Schmidhalter (2015) described that the vertical mode's sensitivity at a coil distance of 1 m is most significant at nearly 0.4 m under the device, while the horizontal mode's sensitivity is highest at a depth of 0.2 m under the device. The complicated correlations between soil ECa, terrain features, and soil properties hamper soil ECa data analysis for mapping the target variable (Lu et al., 2017).

4.2. EMI integration approach

The comparisons of the EMI integration approaches showed strong differences in performances across the three approaches with only a significative improvement of ECe prediction performances when EMI measurements are integrated through the regression co-kriging approach.

Although a larger number of sites were used for calibrating the RF algorithms in approach 1, we did not observe an improvement of the results as observed by Somarathna et al., 2017, Wadoux et al., 2019a, Lagacherie et al., 2020 and Styc et al., 2021. Conversely, the introduction of pseudo values of ECe derived from EMI measurements decreased the performances, which revealed the sensitivity of RF calibration to the uncertainty of data inputs. Alternate models that better account for such uncertainty (Wadoux et al., 2019b) should be applied for improving these results.

The results of approach 2 revealed that using a new covariate obtained from EM38MK2 measurements did not improve significantly the predictions of ECe. Already, Taghizadeh-Mehrjardi et al., 2014 and Wang et al., 2021 investigated soil salinity variation via a regression tree analysis and RF algorithms respectively. Contrary to our result, they emphasized the importance of EM38 data. Taghizadeh-Mehrjardi et al., (2014) performed regression kriging to map ECa data using cubist (regression tree) and kriging with local variograms of residuals to model the deterministic spatial trend and stochastic variation of the spatial model. Although an increase in performance was observed by the residual analysis, estimation performances were still biased according to the reported results. Wang et al., (2021) created ECa map using RF algorithms through environmental variables and electrical magnetic induction data. Then, to study the ECa data's influence on EC prediction, all environmental covariates, including and excluding ECa, were used to generate the EC prediction model. In comparison to our study, regarding the impact of the good correlation between ECe and EM data on the one hand, and sensitivity of the uncertainty indicators (e.g., R2, ME) to the size and the positions of the soil measurements utilized for determining them (Lagacherie et al., 2019) on the other hand, might reveal why different results have been achieved. To avoid these, we used the same test sets in parallel to provide ECa and ECe maps. However, to investigate the models results other influential factors such as spatial density, the range of soil measurements and the environmental covariates need to be considered.

For all the soil layers, a significant increase in performance was observed for approach 3 using regression co-kriging, especially for subsurface soil properties for which the remote sensing data were less appropriate. Taghizadeh-Mehrjardi et al., (2014) reported more reliable predictions in the soil surface layer than the subsoil layer related to the soil's moisture condition. Because the lower conductivity in the soil with a lower moisture content results in restricted penetration ability of EMI signals (Wang et al., 2021) and the accurate EMI data will be achieved when the soil profile contains near to field capacity water content (Corwin and Lesch., 2013; Corwin and Scudiero., 2016). Besides, the more reliable results that have been observed in the subsurface layers might be relevant to this fact that the response of EM38 is affected by various indirect factors, including soil type and texture (Corwin and Scudiero.,2016) as clay content in some part of the basin exceeds 35% in the subsurface horizon of soils (Abtahi,1980; Khormali et al.,2003). Concerning the undefined range of adequate water contents in the previous studies (Corwin and Lesch., 2013; Triantafilis et al., 2001; Moghadas et al., 2016), and the mean and Q3 of the ratio of soil moisture to the water content at the field capacity, these results might be due to the more suitable condition for ECa surveys in the subsurface layer compared to the topsoil layer.

RPIQ, a dimensionless metric that represent the population spread (Bellon-Maurel et al., 2010), revealed the superiority of the third approach better than RMSE, which is related to the dependency of RMSE to the observed data's range (Aman et al., 2015). The RMSE values for the topsoil, subsurface, and subsoil layers were, respectively, 6.26, 6.61, 7.72 dS m−1, that mainly due to the wide range of soil salinity in the study region and the smaller sample size in the extremely saline soils such as lowland and alluvial plains (Wu et al., 2018; Wang et al., 2020). This result was comparable with Koganti et al. (2017), who obtained an RMSE value of 8.31 dS m−1, in the region that the overall range in ECe was 111 dS m−1, and Taghizadeh-Mehrjardi et al. (2014), who reported the RMSE value ranged between 37.5 and 38.4 dS m−1, which is related to the wide range of ECe (244.4 and 237.3 respectively). Furthermore, Wang et al., (2020) and Zare et al., (2015) achieved RMSE values of 6.46 and 5.28 dS m−1 in the region that the range of soil salinity varies, respectively from 0.15 to 77.90 and from 1.9 to 70.3 dS m−1.

The combined effects of the correlation between estimated and predicted values by the QRF model and incorporation regression co-kriging on residuals by considering the uncertainties and bias of the first approach, result in more precise prediction in comparison with the other methods. Coupling regression co-kriging on residuals revealed the effect of the first and third quantiles of data especially in the topsoil (0–0.3 m) and subsurface (0.3–0.6 m) soil ECe with lower median value than mean value.

4.3. Effect of different sample set sizes

Our results clearly showed that the performances of our DSM approach were strongly affected by the size of the calibration data sets. These results confirmed the previous finding of Lagacherie et al., (2020) who verified that the average spacing, strongly influenced the results of a DSM approach, and of Somarathna et al. (2017) and Wadoux et al. (2019a) who, regardless of the algorithms applied to make the DSM models, reported that increasing the amount of input data results in the better performances of Soil Carbon Mapping.

However, our results demonstrated also that using EM38MK2 data was a solution for partially mitigating the sparsity of costly measurements of soil salinity. The gain of performance obtained by integrating EM38MK2 increased as the sizes of ECe measurements decreased. Therefore, by raising the number of measured sites, EM38MK2 data can be a valuable input for broader scale digital soil mapping of ECe where measurements possibilities are much more limited than for this case study. Lagacherie and Gomez (2018), reported that using the VNS-I estimates instead of costly laboratory analysis will be a good decision in the areas where denser spatial sampling is essential for covering the variations.

4.4. Insights on soil salinity distribution

Most saline soils are located in the lake bankside and central parts of the basin that Quaternary sediments with different degrees of salinity make the substratum. Due to the high solubility of halite minerals in salt domes (Hormuz salt formation) and also evaporite formations such as the Sachun, these formations could probably be the major potential source of soil salinity in the study area. Previous studies indicated that the poor quality of the groundwater is mainly relevant to the salt domes and, to a lower degree, from evaporitic and argillitic units (Raeisi et al., 1996). In addition, Raeisi et al.,(1996) and Samani and Gohari (2001) reported that the general flow direction in the Sarvestan basin is downdip from southeast to northwest (from the plain to the Maharlu Lake). Besides, Abtahi., 1980 demonstrated that intensive evaporation from the saline water table, could be a potential source of soil salinity. Therefore, it can be inferred that salt domes and gypsum layers through the runoff and seepage affect groundwater quality and surface deposits, eventually extend the soil salinity in the study region, especially with regard to the intensive conditions throughout recent years, including droughts, an increase in demand for water resources as well as the excessive use of chemical fertilizers.

5. Conclusions

The main lessons of this research works are as follows:

- EM38MK2 could be used in DSM as a surrogate input data for mapping soil salinity
- The selection of an appropriate method for integrating such new input is crucial. Regression co-kriging seems to be the best method to do so.
- The impact of EM38MK2 data on the gains of performance is become greater and greater as the sizes of real measurements of soil salinity decrease.

The present study's contribution is the development of a method for mapping electrical conductivities based on merging the sites with EM38MK2 data and its processing products, in situ ECe data and spatially exhaustive covariates which have not been considered generally for DSM studies. Three different approaches are tested for putting in synergy real measurement and EM38MK2 data. The developed methods suggest that EM38MK2 products could be coupled to enhance the accuracy of DSM outputs, especially where the remote sensing data were less relevant. Hence, in other areas worldwide that soil sensing as alternative data is accessible, this research's future utilization could be possible as a promising way to tackle one of the essential constraints of DSM. The correlations between measured and predicted values and, using regression cokriging on residuals, were the main reasons for the best-proposed method's capability, comparing to the other approaches.

Comparing the models' accuracy with different dataset sizes revealed that the model's prediction accuracy could increase with increasing the sample set's size. According to soil analysis's financial expenditure, increasing sample size in the EM38MK2 survey is an appropriate way for covering the variation of the target variables, especially when there is a lack of intensive field measurements.

Digital soil mapping presents critical information for practical soil rehabilitation programs, policy-making, and natural resources managing. Here, the extended method is simple and clear to reclaim using cheap EM38MK2 data and freely available remote sensing images from its online sources. However, EM38 survey in arid conditions or shallow soils above bedrock is especially problematic because conductance through the liquid pathway reduced when there is insufficient moisture through the depth of investigation. Other soil sensing such as different proximal soil sensing data, remote sensing images, and even unmanned aerial vehicles' images (drone) are suggested as a promising alternative to direct soil measurements that could provide much denser spatial samplings, under some measurement conditions.

Acknowledgements

The authors would like to appreciate Shiraz University and LISAH (INRAE, IRD, Montpellier SupAgro) for providing research facilities.

References

Abbaszadeh Afshar, F., Ayoubi, S., Asghar Besalatpour, A., Khademi, H., Castrignano, A., 2016. Integrating auxiliary data and geophysical techniques for the estimation of soil clay content using CHAID algorithm. J. Appl. Geophys. 126, 87–97.

Abtahi, A., 1980. Soil genesis as affected by topography and time in highly calcareous parent materials under semiarid conditions in Iran. Soil Sci. Soc. Am. J. 44, 329-336.

Aman, S., Simmhan, Y., Prasanna, V.K., 2015. Holistic measures for evaluating prediction models in smart grids. IEEE Trans Knowl Data Eng. 27 (2), 475–488 .

Amezketa, E., de Lersundi, J.D.V., 2008. Soil classification and salinity mapping for determining restoration potential of cropped riparian areas. Land Degrad. Dev. 19(2), 153-164.

Arsoy, S., Ozgur, M., Keskin, E., Yilmaz, C., 2013. Enhancing TDR based water content measurements by ANN in sandy soils. Geoderma 195e196, 133-144.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC - Trends Anal. Chem. 29, 1073–1081.

Ben-Dor, E., Heller, D., Chudnovsky, A., 2008. A novel method of classifying soil profiles in the field using optical means. Soil Sci. Soc. Am. J. 72, 1113-1123.

Bittelli, M., Salvatorelli, F., Pisa, P.R., 2008. Correction of TDR-based soil water content measurements in conductive soils. Geoderma 143, 133-142.

Breiman, L., 2001. Random forests. Mach. Learn. 45(1), 5–32.

Buchanan, S., Triantafilis, J., Odeh, I.O.A ., Subansinghe, R., 2012. Digital soil mapping of compositional particle-size fractions using proximal and remotely sensed ancillary data. Geophysics. 77(4), WB201-WB211.

Carranza, E.J.M., Hale, M., 2002. Mineral imaging with Landsat Thematic Mapper data for hydrothermal alteration mapping in heavily vegetated terrane. Int. J. Remote Sens. 23(22), 4827-4852.

Cheng, Y.B., Ustin, S.L., Riaño, D., Vanderbilt, V.C., 2008. Water content estimation from hyperspectral images and MODIS indexes in Southeastern Arizona. Remote Sens. Environ. 112(2), 363-374.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4, Geosci. Model Dev. 8, 1991–2007.

Cooley, T., Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Chetwynd, J.H., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A .,Bernstein, L.S., 2002, FLAASH, a MODTRAN4-based atmospheric correction algorithm, its application and validation. In Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS), vol. 3. IEEE, 1414–1418.

Corwin, D.L., Lesch, S.M., 2005. Characterizing soil spatial variability with apparent soil electrical conductivity: II. Case study. Comput. Electron. Agri. 46, 135-152.

Corwin, D.L., Lesch, S.M., Shouse, P.J., Soppe, R., Ayars, J.E., 2003. Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity. Agron. J. 95 (2), 352–364.

Corwin, D. L., Lesch, S.M., 2013. Protocols and guidelines for field-scale measurement of soil salinity distribution with ECa–directed soil sampling. J. Environ. Eng. Geophys. 18:1–25. doi:10.2113/JEEG18.1.

Corwin, D.L., Scudiero, E., 2016. Field-scale apparent soil electrical conductivity. Methods Soil Anal. 1(1). doi:10.2136/methods-soil.2015.0038.

Crist, E.P., Cicone, R.C., 1984. A Physically-Based Transformation of Thematic Mapper Data---The TM Tasseled Cap. Geosci. Remote Sens., IEEE Trans. GE-22, 256–263.

Dane, J.H., Hopmans, J.W., 2002. Pressure plate extractor. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis. Part 4. Physical Methods. SSSA Book Ser, 5. SSSA, Madison, WI, pp. 688–690.

Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Hegde, R., Singh, S., Lagacherie, P., 2020, Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. Geoderma Reg. 20, e00250.

Ding, J., Yu, D., 2014. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments. Geoderma 235–36: 316–22. doi:10.1016/j.geoderma.2014.07.028

Dobarco, M.R., Bourennane, H., Arrouays, D., Saby, N.P., Cousin, I., Martin, M.P., 2019. Uncertainty assessment of GlobalSoilMap soil available water capacity products: A French case study. Geoderma 344, 14-30.

Douaoui, A.E.K., Nicolas, H., Walter, C., 2006. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. Geoderma 134, 217–230.

Fallah Shamsi, S.R., Zare, S., Abtahi, S.A., 2013. Soil salinity characteristics using moderate resolution imaging spectro-radiometer (MODIS) images and statistical analysis. Arch. Agron Soil Sci. 59 (4), 471–489.

Feikema, P.M., Baker, T.G., 2011. Effect of soil salinity on growth of irrigated plantation Eucalyptus in south-eastern Australia. Agric. Water Manage. 98:1180–1188. doi:10.1016/j.agwat.2011.03.005

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. 218, 89–103.

Geonics Limited., 2009. EM38–MK2 ground conductivity meter operating manual. Geonics, Ontario, Canada.

Gomez, C., Coulouma, G., 2018. Importance of the spatial extent for using soil properties estimated by laboratory VNIR/SWIR spectroscopy: Examples of the clay and calcium carbonate content. Geoderma 330, 244–253.

Gomez, C., Oltra-Carrio, R., Bacha, S., Lagacherie, P., Briottet, X., 2015. Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery. Remote Sens. Environ. 164, 1–15.

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma 146, 403–411.

Grunwald, S. (Ed.), 2006. Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. CRC Press, New York.

Gyamerah, S.A., Ngare, P., Ikpe, D., 2020. Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function. Agric. For. Meteorol. 280,107808.

Haralick, R., Shanmugan, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics 3, 6, 610-621.

Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987. Image analysis using mathematical morphology. IEEE Trans. Pattern Anal. Mach.Intell. 9 (4), 532-550.

Heil, K., Schmidhalter, U., 2015. Comparison of the EM38 and EM38-MK2 electromagnetic induction-based sensors for spatial soil analysis at field scale. Comput. Electron. Agric. 110, 267–280. https://doi.org/10.1016/j.compag.2014.11.014.

Hengl, T., Heuvelink, G., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput.Geosci. 33 (10), 1301–1315.

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial predictionof soil variables based on regression-kriging, Geoderma 120, 75–93.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B. and Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. Peer J, 6.e5518.

Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Dupuy, J.M., 2011. Combining geostatistical models and remotely sensed data to improve tropical tree richness mapping. Ecol. Indic. 11(5), 1046-1056.

Herrero, J., Hudnall, W.H., 2014. Measurement of soil salinity using electromagnetic induction in a paddy with a densic pan and shallow water table. Paddy Water Environ. 12(1), 263-274.

Hong, Y., Chen, S., Chen, Y., Linderman, M., Mouazen, A.M., Liu, Y., Guo, L., Yu, L., Liu, Y., Cheng, H., 2020. Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon: feature selection coupled with random forest. Soil Tillage Res. 199, 104589.

Huang, B.F., Boutros, P.C., 2016. The parameter sensitivity of random forests. BMC bioinformatics. 17(1), 331.

Huang, C., Wylie, B., Yang, L., Homer, C., Zylstra, G., 2002. Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. Int. J. Remote Sens. 23(8),1741-1748.

Huang, J., Wong,V.N.L., Triantafilis, J., 2014. Mapping soil salinity and pH across an estuarine and alluvial plain using electromagnetic and digital elevation model data. Soil Use Manage. 30, 394–402.

Huete, A.R., 1988. A Soil Adjusted Vegetation Index (SAVI). Remote Sens. Environ. 25 (3), 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83(1-2), 195-213.

Hutter, F., Hoos, H.H., Leyton-Brown, K., 2011. Sequential model-based optimization for general algorithm configuration, 507–523. Berlin, Heidelberg: Springer Berlin Heidelberg.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-boxfunctions. J. Global Optim. 13, 455–492

Karnieli, A., Kaufman, Y.J., Remer, L., Wald, A., 2001. AFRI – aerosol free vegetation index. Rem. Sens. Environ. 77, 10–21.

Khaledian, Y., & Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model.81, 401-418.

Khan, N.M., Rastoskuev, V.V., Shalina, E.V., Sato, Y., 2001.Mapping salt-affected soils using remote sensing indicators-a simple approach with the use of GIS IDRISI, in: 22nd Asian Conference on Remote Sensing. 9.

Khormali, F., Abtahi, A., Mahmoodi, S., Stoops, G., 2003. Argillic horizon development in calcareous soils of arid and semi-arid regions of southern Iran. Catena 53, 273–301.

Khongnawang, T., Zare, E., Zhao, D., Srihabun, P., Triantafilis, J., 2019. Three-Dimensional Mapping of Clay and Cation Exchange Capacity of Sandy and Infertile Soil Using EM38 and Inversion Software.Sensors 19 (18), 3936.

Koganti, T., Narjary, B., Zare, E., Pathan, A.L., Huang, J., Triantafilis, J., 2018. Quantitative mapping of soil salinity using the DUALEM-21S instrument and EM inversion software. Land Degrad. Dev. 29, 1768–1781.

Koyama, C.N., Liu, H., Takahashi, K., Shimada, M., Watanabe, M., Khuut, T., Sato, M., 2017. In-situ measurement of soil permittivity at various depths for the calibration and validation of low-frequency SAR soil moisture models by using GPR. Remote Sens. 9, 1–14. https://doi.org/10.3390/rs9060580.

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied linear statistical models. 5th edition. McGraw Hill, Irwin, New York.

Lagacherie, P., Baret, F., Feret, J.B., Netto, J.M., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sens. Environ. 112 (3), 825–835.

Lagacherie, P., Gomez, C., 2018. Vis-NIR-SWIR Remote Sensing Products as New Soil Data for Digital Soil Mapping, in: McBratney, A.B., Minasny, B. and Stockmann, U. (Eds.), Pedometrics., Springer, pp. 415-437.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., Saby, N.P.A., 2019. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337, 1320–1328.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Nkuba-Kasanda, L., 2020. Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. Geoderma 375, 114503.

Lesch, S.M., Herrero, J., Rhoades, J.D., 1998. Monitoring for temporal changes in soil salinity using electromagnetic induction techniques. Soil Sci. Soc. Am. J. 62 (1) 232–242.

Li, B., Yeh, T.C.J., 1999. Co-kriging estimation of the conductivity field under variably saturated flow conditions. Water Res. Res., 35(12), 3663–3674

Li, N., Zare, E., Huang, J., Triantafilis, J., 2018. Mapping soil cation-exchange capacity using Bayesian modeling and proximal sensors at the field scale. Soil Sci. Soc. Am. J. 82(5), 1203-1216.

Liu, F., Zhang, G.L., Song, X., Li, D., Zhao, Y., Yang, J., Wu, H., Yang, F., 2020. High-resolution and three-dimensional mapping of soil texture of China. Geoderma 361,114061.

Lu, Y., Song, W., Lu, J., Wang, X., Tan, Y., 2017. An examination of soil moisture estimation using ground penetrating radar in dessert steppe. Water (Switzerland) 521, 1–11. https://doi.org/10.3390/w9070521.

Lu, C., Zhou, Z., Zhu, Q., Lai, X., Liao, K., 2017. Using residual analysis in electromagnetic induction data interpretation to improve the prediction of soil properties. Catena 149,176-184.

Maffei, C., Leone, A.P., Vella, M., Meoli, G., Tosca, M., Menenti, M., 2007. Retrieval of vegetation moisture indicators for dynamic fire risk assessment with simulated MODIS radiance. In: Geoscience and Remote Sensing Symposium. IGARSS 2007 IEEE International, pp. 4648–4651.

Martini, E., Comina, C., Priori, S. and Costantini, E.A.C., 2013. A combined geophysical-pedological approach for precision viticulture in the Chianti hills. Boll. Geof. Teor. Appl. 54, 165-181.

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Metternicht, G., Zinck, J.A., 2008. Remote Sensing of Soil Salinization: Impact on Land Management. CRC Press, Taylor and Francis, New York.

Minasny, B., McBratney, A.B., 2006. A conditioned latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388.

Minasny, B., Tranter, G., McBratney, A.B., Brough, D.M., Murphy, B.W., 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. Geoderma 153(1-2),155-162.

Moghadas, D., Taghizadeh-Mehrjardi, R., Triantafilis, J., 2016. Probabilistic inversion of EM38 data for 3D soil mapping in central Iran. Geoderma Reg. 7(2), 230-238.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T., 2011. The use of remote sensing in soil and terrain mapping - A review. Geoderma 162 (1-2), 1–19.

National Cartographic Center of Iran, 2014. Research Institute of National Cartographic Center, Tehran, Iran. http://www.ncc.org.ir

Nouri, M., Gomez, C., Gorretta, N., Roger, J.M., 2017. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring a laboratory regression model. Geoderma 298, 54-66.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E., Papritz, A. J., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil. 4(1),1-22.

Pearson, R.L., Miller, L.D., 1972. Remote mapping of standing crop biomass for estimation of the productivity of the short-grass Prairie, Pawnee National Grasslands, Colorado. In: Proceedings of the 8th International Symposium on Remote Sensing of Environment. ERIM, Ann Arbor (MI), pp. 1357–1381.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691.

Probst, P., Wright, M., Boulesteix, A., 2018. Hyperparameters and Tuning Strategies for Random Forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1–19.

Raeisi, E., Jehbez, O., Moore, F., 1996. Hydrochemical behavior of karstic and evaporitic formations surrounding Sarvestan Plain, Iran. Theor. Appl. Karstol. 9, 165–174.

Rhoades, J.D., Shouse, P.J., Alves, W.J., Manteghi, N.A., Lesch, S.M., 1990. Determining soil salinity from soil electrical conductivity using different models and estimates. Soil Sci. Soc. Am. J. 54, 46–54.

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. NASA/GSFC Type III Final Report, Greenbelt, Md, p. 371.

Richards, L.A., 1954. Diagnosis and improvement of saline and alkali soils. USDA Agriculture Handbook, p. 60.

Slavich, P.G. 1990. Determining ECa-depth profiles from electromagnetic induction measurements. Aust. J. Soil Res. 28, 453–63. doi:10.1071/sr9900443

Samani, N., Gouhari, M., 2001. Hydrogeological Evaluation and Management of Sarvestan Basin, by UNGW Model. J. Sci. I. R. Iran. 12 (1), 37-48.

Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. Soil Sci. Soc. Am. J. 81 (6), 1413–1426.

Spadoni, M., Voltaggio, M., 2013. Contribution of gamma ground spectrometry to the textural characterization and mapping of floodplain sediments. J. Geochem. Explor. 125, 20-33.

Stow, D., Niphadkar, M., Kaiser, J., 2005. MODIS-derived visible atmospherically resistant index for monitoring chaparral moisture content. Int. J. Rem. Sens. 263867–387.

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC bioinformatics, 8(1), 25.

Styc, Q., Gontard, F., Lagacherie, P., 2020. Harvesting spatially dense legacy soil datasets for digital soil mapping of available water capacity in Southern France. Geoderma Reg. e00353.

Szatmári, G., Pásztor, L., 2019 Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. Geoderma 337, 1329–1340.

Taghizadeh-Mehrjardi, R., Ayoubi, S., Namazi, Z., Malone, B.P., Zolfaghari, A.A., Sadrabadi, F.R., 2016. Prediction of soil surface salinity in arid region of central Iran using auxiliary variables and genetic programming. Arid. Land Res. Manag. 30(1),49-64.

Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone. B.P., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma 213, 15–28. doi:10.1016/j.geoderma. 2013.07.020

Taylor, J.R., 1997. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd ed. University Science Books, Sausalito, CA.

Tosti, F., Patriarca, C., Slob, E., Benedetto, A., Lambot, S., 2013. Clay content evaluation in soils through GPR signal processing. J. Appl. Geophys. 97, 69–80.

Triantafilis, J., Buchanan, S.M., 2010. Mapping the spatial distribution of subsurface saline material in the Darling River valley. J. Appl. Geophys. 70(2), 144-160.

Triantafilis, J., Laslett, G.M., McBratney, A.B., 2000. Calibrating and electromagnetic induction instrument to measure salinity in soil under irrigated cotton. Soil Sci. Soc. Am. J. 64, 1009– 1017.

Triantafilis, J., Odeh, I.O.A., Mcbratney, A.B., 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65, 869–878.

Triantafilis, J., Gibbs, I., Earl, N., 2013. Digital soil pattern recognition in the lower Namoi valley using numerical clustering of gamma-ray spectrometry data. Geoderma 192, 407-421.

Triantafilis, J., Lesch, S.M., 2005. Mapping clay content variation using electromagnetic induction techniques. Comput Electron Agr. 46(1-3), 203-237.

Tucker, C., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sens. Environ. 8 , 127–150.

Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Reg. 4, 20–30.

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64.

Viscarra Rossel, R.A., Cattle, S., Ortega, A., Fouad, Y. 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150, 253–266.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. Adv. Agron. 113, 243–291.

Wadoux, A.M.C., Brus, D.J., Heuvelink, G.B.M., 2019a. Sampling design optimization for soil mapping with random forest. Geoderma 355, 113913.

Wadoux, A. M. J., Padarian, J., Minasny, B., 2019b. Multi-source data integration for soil mapping using deep learning. Soil. 5, 107–119.

Wackernagel, H., 1995. Multivariate Geostatistics. Springer Verlag Editions.

Walker, E., Monestiez, P., Gomez, C., Lagacherie, P., 2017. Combining measured sites, soilscapes map and soil sensing for mapping soil properties of a region. Geoderma 300, 64-73.

Wang, F., Yang, S., Wei, Y., Shi, Q., Ding, J., 2021. Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: A case study in Tarim River Basin of southern Xinjiang, China. Sci Total Environ. 754, 142030.

Wang, J., Ding, J., Yu, D., Teng, D., He, B., Chen, X., Ge, X., Zhang, Z.,Wang, Y., Yang, X., Shi, T., Su, F., 2020. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: a comparison between Landsat-8 OLI and Sentinel-2 MSI. Sci. Total Environ. 707, 136092.

Weihermuller, L., Huisman, J.A., Lambot, S., Herbst, M., Vereecken, H., 2007. Mapping the spatial variation of soil water content at the field scale with different ground penetrating radar techniques. J. Hydrol. 340, 205-216.

Wright, M.N., Ziegler, A., 2017. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C ++ and R. J. Stat. Softw. 77(1), 1-17.

Wu, W., Zucca, C., Muhaimeed, A.S., Al-Shafie, W.M., Fadhil Al-Quraishi, A.M., Nangia, V., Zhu, M., Liu, G., 2018. Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. Land Degrad. Dev. 29 (11), 4005–4014.

Yao, R.J., Yang, J.S., Zhao, X., Chen, X., Han, J., Li, X., Liu, M., Shao, H., 2012. A new soil sampling design in coastal saline region using EM38 and VQT method. Clean: Soil Air Water. 40 (9), 972–979.

Yao, R. and Yang, J., 2010. Quantitative evaluation of soil salinity and its spatial distribution using electromagnetic induction method. Agric Water Manage. 97(12), 1961-1970.

Zare, E., Huang, J., Ahmed, M.F., Malik, R.S., Subasinghe, R., Triantafilis, J., 2018. Comparing traditional and digital soil mapping at the district scale using REML analysis. Soil Res. 56, 535–547.

Zare, E., Huang, J., Santos, F.A., Triantafilis, J., 2015. Mapping salinity in three dimensions using a DUALEM-421 and electromagnetic inversion software. Soil Sci. Soc. Am. J. 79(6), 1729-1740.

Zare, S., Fallah Shamsi, S.R., Abtahi, S.A., 2019. Weakly-coupled geo-statistical mapping of soil salinity to Stepwise Multiple Linear Regression of MODIS spectral image products. J. Afr. Earth Sci. 152, 101–114.

Zhang, Y., Ji, W., Saurette, D.D., Easher, T.H., Li, H., Shi, Z., Adamchuk, V.I. and Biswas, A., 2020. Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging. Geoderma 366, 114253.

Zhao, D., Li, N., Zare, E., Wang, J., Triantafilis, J., 2020. Mapping cation exchange capacity using a quasi-3d joint inversion of EM38 and EM31 data. Soil Tillage Res. 200, 104618.