Nutr Cycl Agroecosyst
https://doi.org/10.1007/s10705-023-10303-y

ORIGINAL ARTICLE

# Spatial prediction of soil micronutrients using machine learning algorithms integrated with multiple digital covariates

**Ali Keshavarzi · Fuat Kaya · Levent Başayiğit · Yeboah Gyasi-Agyei · Jesús Rodrigo-Comino · Andrés Caballero-Calvo**

Received: 29 August 2022 / Accepted: 21 July 2023


**Abstract** The design and application of multiple tools to map soil micronutrients is key to efficient land management. While collecting a representative number of soil samples is time consuming and expensive, digital soil mapping could provide maps of soil properties fast and reliably. The objective of this research was to predict the spatial distribution of soil micronutrients within the piedmont plain in northeastern Iran using random forest (RF) and support vector regression (SVR) algorithms. Sixty-eight locations with different land uses were sampled to determine the content of iron, manganese, zinc and copper in the topsoil (0–20 cm). Forty-one digital covariates were used as input to the models and were derived from a digital elevation model, open-source remote sensing (RS) data (Landsat 8 OLI and Sentinel 2A MSI images), WorldClim climate database and maps of soil properties. Covariates were grouped into 11 scenarios: I–III, based on RS data; IV–VI, including RS, topographic, climate and soil covariates; VII, VIII and IX, based only on topographic, climate and soil covariates, respectively; X and XI, based on recursive feature elimination and expert opinion, respectively. The RF algorithm gave 91, 94, 91 and 108% normalized root mean squared error values for iron, manganese, zinc and copper, respectively, for the validation dataset with scenario XI. The most important digital covariates for micronutrients prediction with both RF and SVR models were precipitation seasonality, mean annual temperature and the mean saturation index based on Sentinel 2A MSI data. Digital maps produced at 30 m spatial resolution using scenario XI could be used to effectively identify micronutrient deficiencies and excess hotspots.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10705-023-10303-y.

A. Keshavarzi (✉)
Laboratory of Remote Sensing and GIS, Department of Soil Science, University of Tehran, P.O. Box 4111, Karaj 31587-77871, Iran
e-mail: alikeshavarzi@ut.ac.ir

F. Kaya (✉) · L. Başayiğit
Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Isparta University of Applied Sciences, 32260 Isparta, Çünür, Türkiye
e-mail: fuatkaya@isparta.edu.tr

L. Başayiğit
e-mail: leventbasayigit@isparta.edu.tr

Y. Gyasi-Agyei
School of Engineering and Built Environment, Griffith University, Nathan, QLD, Australia
e-mail: y.gyasi-agyei@griffith.edu.au

J. Rodrigo-Comino · A. Caballero-Calvo
Departamento de Análisis Geográfico Regional y Geografía Física, Facultad de Filosofía y Letras, Campus Universitario de Cartuja, Universidad de Granada, 18071 Granada, Spain
e-mail: jesusrc@ugr.es

A. Caballero-Calvo
e-mail: andrescaballero@ugr.es

Published online: 13 August 2023

**Keywords** Arid regions · Digital soil mapping · Expert opinion · Feature selection · Remote sensing

## Introduction

Soil nutrients are key factors for crop production and play important roles in plant growth but can limit production where insufficient amounts prevail (Miran et al. 2021; Sharma et al. 2022). Soil nutrients content is susceptible to modification by human activities, and is therefore difficult to be assessed, particularly at the catchment scale or even at the hillslope scale (Wang et al. 2008). The essential nutrients, without which plant growth will not be possible, come from the mineral components of the soil. Depending on the amount of nutrients needed by the plant, the distinction can usually be made between macronutrients and micronutrients (Blume et al. 2016a). The determination of soil nutrient contents, both micronutrients and macronutrients, is indispensable as sustainable agricultural production requires optimum levels (Kaya and Başayiğit 2022). The real pressure of micronutrient deficiencies in arable land will be on the achievement of optimum crop growth and yield, but also on human populations with high dietary dependence on crops grown in these areas.

Regarding human health, the levels of micronutrients should be at the optimum as both deficiency and excess have negative effects. Deficiency of micronutrients in humans can cause serious diseases such as anemia and neutropenia (Denton-Thompson and Sayer 2022).

The spatial variability of micronutrient levels in the topsoil is influenced by the complex interactions between climate, time, soil parent material and physicochemical properties, topography, vegetation, land use, and possibly exogenous inputs from industry (Srisomkiew et al. 2022). Therefore, a better understanding of the dynamic variation of micronutrients and the factors contributing to their levels in the soil is paramount (Zhu et al. 2021).

Developing adequate agricultural management strategies and policies, along with minimising the environmental impact on farming revenue, requires detailed site-specific soil knowledge (Snapp 2022). While measuring the factors affecting soil micronutrient distribution in agricultural areas in arid and semi-arid regions is often difficult, it is crucial to understand the factors influencing their availability and concentration for proper management of soil fertility and land use. The presence of calcium in the form of calcium carbonate ($CaCO_3$) compounds, particularly in arid and semi-arid areas, can result in the conversion of soil micronutrients into an unavailable form (Blume et al. 2016b). Thus, plant-available micronutrient maps are critical in arid regions where both irrigated-rainfed annual crops and perennial fruit plants farming are practiced. It is crucial to develop more precise management practices that provide micronutrients for these crops (Naimi et al. 2021).

Accurately estimating micronutrients content at any scale requires a wide range of methods capable of monitoring the soil surface. Alloway (2013) used the diethylenetriaminepentaacetic acid (DTPA) test to measure the available micronutrients in soil, which refers to the portion of micronutrients that are readily available for plant uptake. Digital soil mapping (DSM), a sub-science of pedometry, has been recently used to map soil characteristics (Lark et al. 2014). The DSM uses data science and predictive modelling approaches to establish the relationships between digital covariates, location-based soil observations and soil properties (Wadoux et al. 2021). Malone et al. (2022) obtained a data matrix based on the intersection of continuous and categorical soil data in a geographic setting with a pedologically meaningful set of environmental layers. Machine learning (ML) models are used to establish linear or nonlinear relationships in data sets to obtain a model equation (Nussbaum 2022). Remote sensing and DSM were employed to study the spatial variability of soil properties and soil nutrients (Miran et al. 2021). The successful integration with DSM has led to active work on global (FAO and ITPS 2018), continental (Hengl et al. 2017) and regional (Zhang et al. 2020; Vasu et al. 2021; Kaya and Başayiğit 2022) scales for the diagnosis of the spatial distribution of soil micronutrients associated with soil fertility. The DSM methods are cost-effective and allow the creation of high spatial resolution ($\leq$30 m) maps characterised by high accuracy. However, there is limited information on the potential use of RS-based estimators of soil nutrients in arid and semi-arid regions (Kaya and Başayiğit 2022). Soil data for spatial analysis is lacking in arid and semi-arid regions, making it difficult for establishing land use management processes and policies (Smith et al. 2019).

In recent times the number of environmental covariates used as input to DSM is increasing but covariate selection to limit the number of input covariates has been highlighted (Liang et al. 2020; Wadoux et al. 2021). Without a doubt, increasing the number of digital covariates improves ML model accuracy, but this approach increases uncertainty in the input data that makes the results difficult to interpret in soil science. Thus, there is a need to balance covariate parsimony and model performance through an appropriate covariate selection on case-based reasoning and pedological relevance (Liang et al. 2021; Wadoux et al. 2020). The abundance of digital data and the use of expert knowledge as a primary driver in the various areas of soil science have several implications for DSM that are yet to be documented. As a matter of fact, "scenario" and "grouping" strategies based on digital surrogates of soil formation factors can be used in evaluating digital covariates, especially in DSM (Wadoux et al. 2021).

Despite the availability of methodological approaches to integrate different types of ML models and digital covariates grouping approaches into the DSM methodology (Hengl et al. 2017; Zhu et al. 2021), the estimation of soil micronutrient content with DSM has rarely been studied.

The specific objectives of the current study were to: (i) construct a digital map (30 m × 30 m grid scale) of micronutrients in the topsoil using two ML algorithms (Random forests and Support vector regression) at the Piedmont site in Iran; (ii) include an expert opinion to overcome the challenge of selecting digital covariates in DSM; (iii) assess the digital maps created by the two ML algorithms using a qualitative soil scientist approach, and (iv) investigate the soil and environmental factors associated with micronutrient variability.

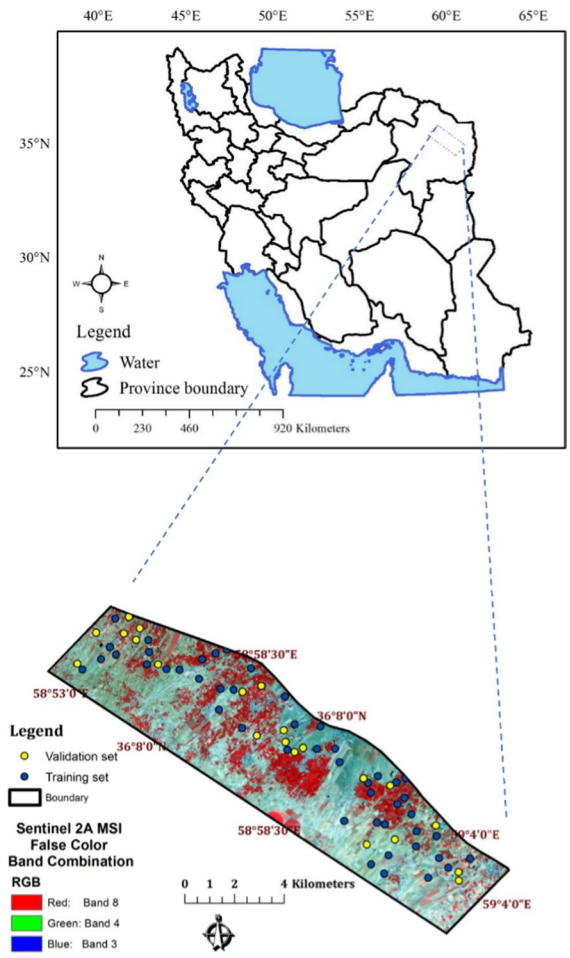



**Fig. 1** Location of the case study area in Northeast Iran (top) and the spatial distribution of the soil samples used for training and validation overlaid on a Sentinel 2A MSI false-colour band combination

## Materials and methods

### Study area

The case study area, which covers ~ 100 km$^2$ in the Neyshabur plain of Khorasan-e-Razavi province, is located in the semi-arid region of Northeast Iran (Fig. 1). It is bounded by latitudes 36 02′ N and 36 08′ N, and longitudes 58 53′ E and 59 04′ E, with an average daily temperature of 14.5 °C and an average annual precipitation of 233.7 mm (Bagherzadeh et al. 2016). The geological composition of the area mainly consists of low levels of Piedmont fan and valley terrace deposits, and the major soil types are Entisols and Aridisols according to the Soil Survey Staff (2014).

2.2. Reference dataset.

A total of 68 sites were selected according to a stratified scheme. A portable handheld Global Positioning System (GPS) device was used to log the coordinates of the sampling sites. Triplicate composite samples were collected for each site surface within the top 0–20 cm layer by a soil auger during the summer 2018. The air-dried soil samples were

crushed and particles with diameters less than 2 mm were obtained by sieving. Soil iron (Fe), copper (Cu), manganese (Mn), and zinc (Zn) were extracted with DTPA solution (0.005 M DTPA + 0.01 M $CaCl_2$ + 0.1 M triethanolamine, pH 7.3) and measured with an atomic absorption spectrophotometer (Perkin Elmer, AAnalyst 700, USA) (Lindsay and Norvell 1978).

## Covariates used for DSM

Forty-one digital covariates that may affect micronutrient content were used (Fig. 2).

### *Topographic and remote-sensing based covariates*

A digital elevation model (DEM) was the source of the topographic digital covariates (ALOS PALSAR 2021). The terrain attributes were calculated from a 30 m × 30 m spatial resolution of the DEM (Table 1). Different sources of RS data provided the vegetation-based and soil-based digital covariates that represent organisms and parent material (Kaya and Başayiğit 2022). Four different years (2018, 2019, 2020, and 2021) of Landsat 8 OLI sensor, Collection 2 Level 2 Science Product, 160 path, 35 Row, and Collection 2 level satellite images were used in the study. The Landsat 8 Surface Reflectance (SR) Science Products were produced using the Land Surface Reflectance Code (LaSRC) specialised software (Sayler and Zanter 2021). The calibration parameters were obtained from the Landsat SR data product provided by Earth Explorer (EE-https://earthexplorer.usgs.gov/). The ArcGIS 10.8-Arctoolbox—Spatial Analyst-Map Algebra-Raster Calculator tool (ESRI 2021) was used to process the selected bands. Furthermore, Sentinel 2A MSI sensor, Level 2A Bottom of Atmosphere, R020 orbit number defining the "Track", T40SFF ID of the area that has been visualised defining the "Granule", indicate the satellite images from three different recent years (2019, 2020, and 2021) that were used (Table 1) (ESA 2015). Mean synthesis bands were produced from the bands of both satellite images of different years, but for the same seasons.

### *Climate and soil-related covariates*

Climate-based digital covariates from the "WorldClim 2" dataset (Fick and Hijmans 2017) are presented in Table 1. The raw covariates which have a spatial resolution of 30 arc-second, were converted to the EPSG:32640 projection system and resampled to 30 m × 30 m spatial resolution.

The spatial distribution maps of relevant soil properties, such as soil particle size fractions and pH, were generated at 30 m × 30 m pixel resolution using geostatistical analysis (ordinary kriging) with the "Geostatistical Wizard" tool in ArcGIS 10.8. (ESRI 2021). These generated raster maps were used as environmental covariates. The mapping process was performed with the spherical semi-variogram model at points that were on average 530 m

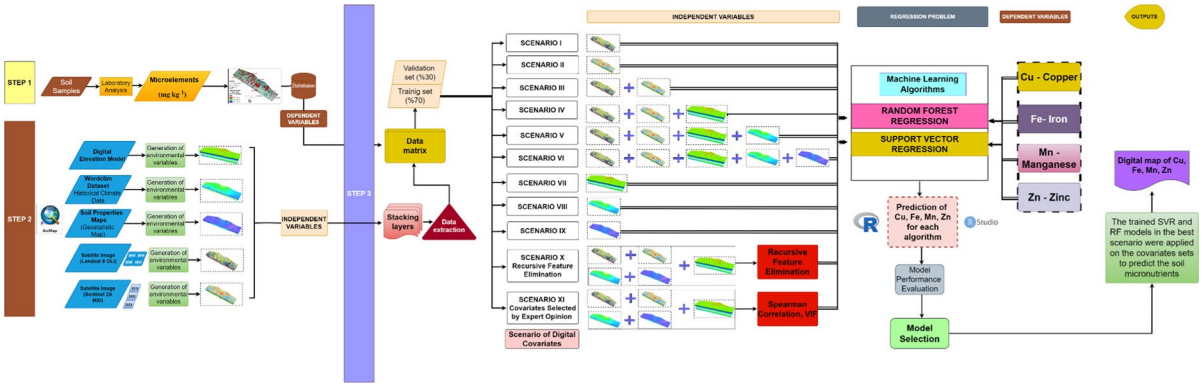



**Fig. 2** Grouping of digital covariates to constitute the scenarios and modelling process (I: RS-Landsat 8 OLI, II: RS-Sentinel 2A MSI, III: RS(All), IV: RS (All) + Topographic, V: RS (All) + Topographic + Climate, VI: RS (All) + Topographic + Climate + Soil, VII: Topographic, VIII: Climate, IX: Soil, X: Recursive Feature Elimination, XI: Covariates selection by expert opinion and the modelling process workflow

**Table 1** List of covariates used in the study

| Auxiliary variables | | Environmental covariate | |
|---|---|---|---|
| Digital elevation model (DEM) | | Elevation (m) | (https://search.asf.alaska.edu/; ALOS PALSAR 2021) |
| | | Slope (%) | Hengl and Reuter (2008) and Gruber and Peckham (2009) |
| | | Profile curvature | |
| | | Planform curvature | |
| | | Topographic wetness index (TWI) | |
| | | Flow Accumulation | |
| | | Stream power index (SPI) | |
| Remote sensing data (RS) | Sentinel 2A MSI[a] | Modified secondary soil-adjusted vegetation index (MSAVI2) (Mponela et al. 2020) | |
| | | Topsoil grain size index (Xiao et al. 2006) | |
| | | Saturation index (Hounkpatin et al. 2018) | |
| | | Normalized clay index (Brown et al. 2017) | |
| | | Normalized difference vegetation index— (NDVI) (Brown et al. 2017) | |
| | | Green normalized difference vegetation index (GNDVI) (Gitelson et al. 1996) | |
| | | Band 2-blue (central wavelength: 490 nm)-10 m | |
| | | Band 3-green (central wavelength: 560 nm)-10 m | |
| | | Band 4-red (central wavelength: 665 nm)-10 m | |
| | | Band 5-vegetation red-edge (central wavelength: 705 nm)-20 m | |
| | | Band 6-vegetation red-edge (central wavelength: 740 nm)-20 m | |
| | | Band 7-vegetation red-edge (central wavelength: 783 nm)-20 m | |
| | | Band 8-near infrared (NIR) (central wavelength: 842 nm)-10 m | |
| | | Band 8a-vegetation red-edge (central wavelength: 865 nm)-20 m | |
| | | Band 11-shortwave infrared (SWIR) (central wavelength: 1610 nm)-20 m | |
| | | Band 12-shortwave infrared (SWIR) (central wavelength: 2190 nm)-20 m | |
| | Landsat 8 OLI[b] | Topsoil grain size index (Xiao et al. 2006) | |
| | | Saturation index (Hounkpatin et al. 2018) | |
| | | Normalized clay index (Brown et al. 2017) | |
| | | Normalized difference vegetation index—(NDVI) (Brown et al. 2017) | |
| | | Green normalized difference vegetation index (GNDVI) (Gitelson et al. 1996) | |
| | | Band 2-blue (450–515 nm)-30 m | |
| | | Band 3-green (525–600 nm)-30 m | |
| | | Band 4-red (630–680 nm)-30 m | |
| | | Band 5-near infrared (NIR) (845–885 nm)-30 m | |
| | | Band 6-SWIR 1 (1560–1660 nm)-30 m | |
| | | Band 7-SWIR 2 (2100–2300 nm)-30 m | |
| Climatic variables (CL) | | BIO-1- mean annual temperature (MAT) (°C) | |
| | | BIO-12-annual precipitation (mm) | |
| | | BIO-15-precipitation seasonality | |
| | | Total solar radiation (kJ $m^{-2}$ $year^{-1}$) | |
| Soil attributes (S) | | Clay map produced with Kriging (%) | |
| | | Sand map produced with Kriging (%) | |
| | | pH map produced with Kriging | |

[a]Sentinel 2A MSI ProductID information

S2A_MSIL2A_20190703T064631_N0212_R020_T40SFF_20190703T112808

S2A_MSIL2A_20200707T064631_N0214_R020_T40SFF_20200707T112619

S2A_MSIL2A_20210831T064631_N0301_R020_T40SFF_20210831T094435

[b]Landsat 8 OLI ProductIDinformation

LC08_L2SP_160035_20180630_20200831_02_T1

LC08_L2SP_160035_20190703_20200827_02_T1

LC08_L2SP_160035_20200721_20200911_02_T1

LC08_L2SP_160035_20210825_20210901_02_T1

apart to calculate the lag size and the average distance between the points. The "Spatial Statistics tools—Analysis model—Average nearest neighbour" tool in ArcGIS was used to determine the lag size value.

*Grouping of the digital covariates*

All covariates were processed to conform to the same areal extent, grid size resolution (30 m), and grid centred using the nearest neighbour resampling function of the ArcGIS 10.8 (ESRI 2021) software. (EPSG:32640) coordinate reference system was used.

The 41 digital covariates were combined into a raster stack, and the R "raster" package (Hijmans 2020) was used to extract the values at the 68 soil sample locations (Fig. 2). Eleven scenarios (Fig. 2) were set as combinations of input digital covariates (RS-based, topographic, climatic and soil) considering the parsimony approach in the modelling process (Wadoux et al. 2021; Azizi et al. 2022). Some scenarios used a single dataset (for example Scenario I use only Landsat 8 OLI-based digital covariates) while others involved the integration of multiple datasets (for example scenario VI have all RS-based, topographic, climate and soil digital covariates). Scenario X consisted of digital covariates selected with the "rfe" (recursive feature elimination) function (Kuhn 2020). In scenario XI, after identifying the reasonable digital covariates according to expert opinion, the Variance Inflation Factor (VIF) was calculated using the MINITAB software (V. 17.1.0) to assess the multicollinearity among covariates. Digital covariates with a VIF value greater than ten were removed. Then, the Spearman rank correlation coefficient was calculated to determine the degree of association between the soil micronutrients and the reasonable digital covariates. Correlations were performed in the R "stats" package (R Core Team 2022). Finally, the expert opinion scenario XI was created by considering the VIF results and the Spearman correlation values (Naimi et al. 2021; Azizi et al. 2022).

The modelling process

The random forest (RF) algorithm (R package "randomForest", Liaw and Wiener 2002), and the support vector regression (SVR) algorithm (R package "e1071", Meyer et al. 2020) were used to predict soil micronutrients by regression, separately for each covariate scenario (Fig. 2). The reference dataset of micronutrient concentration was split into a training set (70%, n = 47) and a validation set (30%, n = 21). The models were trained on the training dataset and their performances were evaluated on the validation dataset.

*Random forest*

The study used the RF method (Breiman 2001), which was an advanced version of bagging that involves constructing many uncorrelated regression trees and combining their predictions. The optimal value of the "mtry" parameter was determined through a parameter tuning process using the "train" function in the "caret" package (Kuhn 2020) in the R Core Environment software (Version 4.2.1). The results of the process are reported in the *GitHub* repository (Kaya 2022). Default values were used for other parameters such as "ntree". The "importance" function of the "randomForest" package was used to assess the relative importance of each digital covariate. The mean square error (MSE) was adopted in the out-of-bag (OOB) analysis of regression models. During the analysis the software calculated two indicators: %IncMSE (increase in mean standard error) and IncNodePurity. The %IncMSE was determined by comparing the MSE values of each regression tree, both with and without the relevant predictor. The mean difference between the with and without the relevant predictor is then normalized by the standard deviation of the differences. IncNodePurity represents the average reduction in node impurity achieved by splitting the predictors during the tree-building process. The node impurity is measured by the residual sum of squares (Breiman 2001).

*Support vector regression*

Support vector machine was developed by Cortes and Vapnik (1995). It is commonly used for classification problems but can also be used for regression problems as demonstrated in Drucker et al. (1997) as well as applied to soil modelling by Pasolli et al. (2011). The relationship between a dependent variable $y$ (soil micronutrient concentration) and independent variables $x$ (the set of digital covariates) is expressed as:

$$y = f(x) + e \quad (1)$$

To handle non-linear relationships between $y$ and $x$, the data were transformed into a higher-dimensional space through kernel techniques. The radial basis function (RBF) kernel, which is a Gaussian function, was used for this purpose (Pasolli et al. 2011). The values of "gamma" and "cost" parameters were initially determined using the repeated tenfold cross-validation option (Kuhn 2020; Meyer et al. 2020) and subsequentially optimised through a parameter searching process (Kaya 2022). Following Yigini et al. (2018), the importance of each digital covariate in the model was calculated.

### Model performance evaluation

The performance statistics of normalised root mean square error (NRMSE) and mean absolute percentage error (MAPE) were calculated and model performances were evaluated in the training and validation sets. The NRMSE is given as (Zambrano-Bigiarini 2020):

$$\mathrm{NRMSE} = 100 \times \left( \frac{\sqrt{\frac{\sum_{i=1}^{n} (O_i - P_i)^2}{n}}}{Normconstant} \right) \quad (2)$$

where n is the sample size, $P_i$ and $O_i$ are the predicted and observed values, respectively, and the *Normconstant* was set to the default value of the standard deviation of the observations. The "nrmse" function in the "hydroGOF" R package (Zambrano-Bigiarini 2020) was used to estimate NRMSE. The MAPE is expressed as a percentage as (Gopp and Savenkov 2019):

$$\mathrm{MAPE} = \left( \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathrm{O}i - \mathrm{P}i|}{\mathrm{O}i} \right) \times 100 \quad (3)$$

The agreement between observed and modelled values using scenario XI was also evaluated with Taylor diagrams which aggregate the correlation coefficient (r), the standard deviation of measured data (SD), and the RMSE (Wadoux et al. 2022). The "TaylorDiagram" function of the R package "openair" (Carslaw and Ropkins 2012) was used. The "scatterPlot" function of R package "openair" with method "hexbin" was used to show the relationship between the observed and the predicted values (Carslaw and Ropkins 2012; Carr et al. 2021).

## Results

### Statistics of soil micronutrients and correlations with digital covariates

Descriptive statistics of micronutrients concentration in the topsoil are shown in Table 2. Slight to moderate alkaline conditions prevailed in the study area. Micronutrients concentration exhibited positive skewness for both the training and validation datasets.

Results highlighted statistically significant ($p < 0.05$) correlation between the four micronutrients concentration and four digital covariates from distinct groups (remote sensing, topography, climate, and soil). Spearman correlation coefficients indicated a significant correlation between Fe and Zn ($r = 0.22$, $p < 0.05$), Clay ($r = -0.28$, $p < 0.05$), Sand ($r = 0.17$, $p < 0.05$), and Elevation ($r = 0.43$, $p < 0.05$) (Figure S1). Manganese was significantly correlated with Cu ($r = 0.34$, $p < 0.05$), pH (r: $-0.35$, $p < 0.05$), BIO-15 (precipitation seasonality) ($r = 0.31$, $p < 0.05$), and solar radiation ($r = 0.29$, $p < 0.05$) (Figure S1). Copper showed a significantly positive correlation with the different spectral bands, but it had a strong negative correlation with BIO-12 ($r = -0.23$, $p < 0.05$), and a positive correlation with BIO-1 ($r = 0.27$, $p < 0.05$) (Figure S1). Zinc demonstrated a significantly negative correlation with the different spectral bands, while it showed a strong positive relationship with the BIO-12 (annual precipitation) ($r = 0.42$, $p < 0.05$) and negative correlation with BIO-1 (mean temperature) ($r = -0.53$, $p < 0.05$) (Figure S1).

### Evaluation of model performances

In the expert opinion scenario (scenario XI), the lowest MAPE value was obtained for Fe, and the highest value for Zn in the validation dataset with both models (Table 3). The lowest NRMSE values were obtained for Fe and Zn and the highest values for Cu in the validation datasets with the RF model for scenario XI (Table 3). Using the same scenario,

**Table 2** Descriptive statistics of measured micronutrients concentration (mg $kg^{-1}$)

| Dataset | Variable | Mean | SD | CV (%) | Min | Median | Max | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|
| All datasets (N = 68) | pH | 7.90 | 0.19 | 2.43 | 7.50 | 7.90 | 8.30 | − 0.01 | − 0.59 |
| | Fe | 2.31 | 0.68 | 29.23 | 1.24 | 2.28 | 4.22 | 0.66 | 0.20 |
| | Mn | 7.18 | 4.06 | 56.51 | 1.64 | 6.04 | 21.06 | 1.61 | 2.53 |
| | Zn | 2.84 | 3.78 | 133.05 | 0.28 | 1.09 | 13.94 | 1.68 | − 0.34 |
| | Cu | 1.16 | 0.28 | 24.81 | 0.72 | 1.13 | 1.94 | 0.47 | − 0.34 |
| Training dataset (N = 47) | Fe | 2.25 | 0.66 | 29.17 | 1.24 | 2.14 | 4.22 | 0.72 | 0.57 |
| | Mn | 7.21 | 4.20 | 58.25 | 2.82 | 5.86 | 21.06 | 1.64 | 2.67 |
| | Zn | 2.69 | 3.60 | 133.82 | 0.3 | 1.06 | 13.9 | 1.71 | 1.74 |
| | Cu | 1.19 | 0.28 | 24.12 | 0.72 | 1.22 | 1.94 | 0.31 | − 0.38 |
| Validation dataset (N = 21) | Fe | 2.47 | 0.71 | 28.98 | 1.4 | 2.42 | 3.94 | 0.56 | − 0.16 |
| | Mn | 7.11 | 3.82 | 53.69 | 1.64 | 6.2 | 18.08 | 1.6 | 2.88 |
| | Zn | 3.18 | 4.23 | 133.04 | 0.28 | 1.42 | 13.94 | 1.69 | 1.6 |
| | Cu | 1.11 | 0.29 | 26.33 | 0.74 | 1.04 | 1.82 | 0.92 | 0.55 |

*N* number of samples, *SD* Standard Deviation, *CV* Coefficient of Variation, *Max.* Maximum, *Min.* Minimum

relatively higher NRMSE values were obtained for Fe, Mn, and Zn while relatively lower values were obtained for Cu in the validation sets with the SVR model compared to the RF model (Table 3). Scenarios based on Landsat 8 OLI (scenario I) and Sentinel 2A MSI (scenario II) satellite images produced similar results in the estimation of the micronutrient contents for both ML algorithms (Table 3). With the RF model, high model accuracy in predicting Fe and Zn was obtained using scenario VI that included soil properties as covariates.

Examination of the Taylor diagrams indicated lower RMSE values for Fe and Cu (Fig. 3a, c) and higher RMSE values for Mn and Zn (Fig. 3b, d). Generally, the ML models performed better with the training datasets compared with the validation datasets for all micronutrients.

## Importance of the environmental covariates for random forests and support vector regression

The findings of this study indicated that the distribution of micronutrients in the study area was largely influenced by climate-based and topography-based digital covariates (Figures S2 and S3). The RF model revealed that BIO1- Mean Annual Temperature (MAT) was the most important predictor of the four micronutrients (Figures S2a, c, e, g) based on the %IncMSE indicator. The same digital covariate emerged as the most important predictor based on the IncNodePurity indicator except for Zn (Figure S2b, d, f, h).

With the SVR modelling, the most important predictors differed among micronutrients (Figure S3a–d). The model results indicated that, while climatic covariates were more important for Cu (Figure S3a) and Mn (Figure S3c), topographic digital covariates were found to be important for Fe (Figure S3b). In the case of Zn, the saturation index produced from Sentinel 2A MSI bands was found to be the most important digital covariate (Figure S3d).

## Spatial prediction of soil micronutrients

The scenario XI was chosen as the best for generating micronutrient maps. Therefore, the models based on this scenario were applied to the raster stack to produce micronutrient maps.

Maps produced by the RF model (Fig. 4a, d, g, j) exhibited spatial heterogeneity of micronutrient concentration while maps based on SVR models (Fig. 4b, e, h, k) were fairly smoothing distributed with poor

**Table 3** Comparisons of the accuracy of the support vector regression (SVR) and random forest (RF) models for the training and the validation datasets (MAPE: %, NRMSE: %)

| Model | Scenario | Training | | Validation | | Training | | Validation | |
|---|---|---|---|---|---|---|---|---|---|
| | | Fe | | | | Mn | | | |
| | | MAPE | NRMSE | MAPE | NRMSE | MAPE | NRMSE | MAPE | NRMSE |
| RF | I | 11 | 49 | 19 | 91 | 28 | 54 | 54 | 108 |
| | II | 12 | 48 | 19 | 91 | 25 | 54 | 58 | 114 |
| | III | 12 | 48 | 18 | 89 | 26 | 55 | 55 | 110 |
| | IV | 12 | 48 | 19 | 89 | 26 | 53 | 55 | 109 |
| | V | 12 | 48 | 18 | 88 | 24 | 51 | 53 | 102 |
| | VI | 12 | 47 | 19 | 88 | 23 | 50 | 51 | 100 |
| | VII | 12 | 49 | 24 | 106 | 28 | 54 | 53 | 103 |
| | VIII | 12 | 47 | 17 | 90 | 21 | 51 | 48 | 94 |
| | IX | 11 | 47 | 24 | 98 | 23 | 51 | 43 | 81 |
| | X | 12 | 47 | 18 | 86 | 23 | 49 | 57 | 104 |
| | XI | 12 | 47 | 19 | 91 | 24 | 50 | 50 | 94 |
| SVR | I | 14 | 73 | 17 | 92 | 11 | 57 | 42 | 96 |
| | II | 20 | 88 | 21 | 97 | 11 | 55 | 44 | 99 |
| | III | 19 | 84 | 21 | 98 | 7 | 10 | 51 | 97 |
| | IV | 21 | 90 | 22 | 103 | 7 | 10 | 52 | 98 |
| | V | 3 | 10 | 21 | 100 | 7 | 10 | 52 | 98 |
| | VI | 6 | 41 | 22 | 100 | 7 | 10 | 52 | 98 |
| | VII | 20 | 85 | 22 | 102 | 34 | 98 | 38 | 101 |
| | VIII | 6 | 30 | 18 | 78 | 25 | 86 | 41 | 96 |
| | IX | 8 | 57 | 18 | 92 | 19 | 65 | 40 | 97 |
| | X | 6 | 41 | 22 | 102 | 7 | 10 | 52 | 98 |
| | XI | 3 | 10 | 21 | 98 | 10 | 54 | 44 | 98 |
| | | Zn | | | | Cu | | | |
| RF | I | 109 | 49 | 178 | 93 | 11 | 50 | 30 | 117 |
| | II | 113 | 51 | 232 | 94 | 10 | 46 | 27 | 116 |
| | III | 116 | 49 | 233 | 91 | 10 | 48 | 27 | 114 |
| | IV | 119 | 49 | 217 | 95 | 10 | 48 | 26 | 109 |
| | V | 119 | 48 | 213 | 92 | 8 | 39 | 27 | 110 |
| | VI | 91 | 38 | 241 | 84 | 8 | 38 | 27 | 112 |
| | VII | 141 | 54 | 206 | 96 | 11 | 53 | 24 | 98 |
| | VIII | 94 | 47 | 221 | 87 | 9 | 45 | 27 | 100 |
| | IX | 68 | 36 | 223 | 79 | 10 | 49 | 27 | 108 |
| | X | 87 | 46 | 304 | 100 | 8 | 37 | 26 | 112 |
| | XI | 118 | 45 | 210 | 91 | 8 | 39 | 26 | 108 |
| SVR | I | 45 | 23 | 218 | 99 | 11 | 63 | 25 | 104 |
| | II | 44 | 10 | 260 | 98 | 3 | 14 | 24 | 101 |
| | III | 44 | 10 | 232 | 96 | 3 | 10 | 24 | 102 |
| | IV | 44 | 10 | 218 | 95 | 5 | 35 | 24 | 101 |
| | V | 44 | 10 | 243 | 97 | 3 | 13 | 24 | 101 |
| | VI | 44 | 10 | 244 | 96 | 3 | 10 | 24 | 102 |
| | VII | 45 | 20 | 218 | 94 | 13 | 73 | 23 | 100 |
| | VIII | 68 | 54 | 155 | 92 | 3 | 17 | 26 | 98 |
| | IX | 43 | 10 | 170 | 84 | 18 | 87 | 26 | 104 |
| | X | 79 | 54 | 263 | 107 | 3 | 10 | 24 | 102 |

**Table 3** (continued)

| Model | Scenario | Training | | Validation | | Training | | Validation | |
|---|---|---|---|---|---|---|---|---|---|
| | | Fe | | | | Mn | | | |
| | | MAPE | NRMSE | MAPE | NRMSE | MAPE | NRMSE | MAPE | NRMSE |
| | XI | 43 | 10 | 234 | 95 | 3 | 10 | 23 | 96 |

I: RS-Landsat 8 OLI, II: RS-Sentinel 24 MSI, III: RS(All), IV: RS (All) + Topographic, V: RS (All) + Topographic + Climate, VI: RS (All) + Topographic + Climate + Soil, VII: Topographic, VIII: Climate, IX: Soil, X: Recursive Feature Elimination, XI: Covariates selection by expert opinion. Abbreviations. *NRMSE (%)* Normalised Root Mean Square Error, *MAPE (%)* Mean Absolute Percentage Error

prediction of extreme values that were limited in the measured dataset. Considering the actual micronutrient concentrations of the sampling sites (Fig. 4, c, f, I, l), the RF algorithm was found to be more useful than the SVR, particularly for areas with relatively higher micronutrient concentrations.

## Discussion

### Variation of micronutrients in the study area

The coefficient of variation of Fe in the dataset was 29.23%, which is considered an average value, a value of 24.81% for Cu is considered moderate, and a value for Mn of 56.51% is considered very variable, according to the limits reported by Wilding (1985) (Table 2). In the studied arid and semi-arid area, land use combined with the parent material and geochemical factors influence the spatial distribution of soil micronutrients (Keshavarzi et al. 2022). The weak correlations of topographical digital covariates (except elevation) with soil micronutrients were due to the fact that the topographical digital covariates exhibited low variability in the study area, as observed by Vasu et al. (2021). All the micronutrients, except Mn, had a significant correlation with Clay, and only Fe did not have a significant correlation with pH (Figure S1). Foroughifar et al. (2013) showed that the distribution of micronutrients in dry farming soils is related to clay content and pH value, which supports our findings.

In accordance to the breakpoints established by Sharma et al. (2022), the levels of Fe ($<4.5$ mg kg$^{-1}$) in the study area are insufficient, while the levels of Cu ($>0.5$ mg kg$^{-1}$), Mn ($>4.0$ mg kg$^{-1}$), and Zn ($>1.2$ mg kg$^{-1}$) in the middle and east of the study area are relatively sufficient, particularly for Zn. This information can be used to address the soil micronutrient deficiencies in the region and make informed decisions about soil fertility management, particularly in areas of permanent crops. It was noted that, and as described by Keshavarzi et al. (2022), the study area has permanent crops located in the central, north-western, and north-eastern regions.

### Model performance

The performance of the models based on the various combinations of the digital covariates varied, but only more complex models with a more extensive set of covariates (e.g., remote sensing, topography, climate, and soil) were able to achieve relatively better performances in the training and validation datasets (Table 3, Table S1). The expert opinion (scenario XI) and RF caused a reduction in NRMSE values for the four micronutrients between 0.11% (Zn) and 8.29% (Mn) compared to the average values of scenarios I-X for the validation dataset (Table S1). The reduction was also observed for MAPE values with the RF model (Table S1). The reduction in the performance statistics values caused by scenario XI compared with the individual 10 scenarios is also quite variable for SVR (Table 3).

As previously shown by Miran et al. (2021), the use of multispectral satellite data for mapping soil micronutrient content may also support the effectiveness of scenario I and scenario II. For all nutrients except Cu, RF yielded marginally better results than SVR. Miran et al. (2021) obtained RMSE values of 1.65 mg kg$^{-1}$ for Fe, 1.66 mg kg$^{-1}$ for Mn, 0.17 mg kg$^{-1}$ for Zn, and 0.23 mg kg$^{-1}$ for Cu using

only Landsat 8 OLI satellite images in dry agricultural areas in north-western Iran. Zhu et al. (2021) used climate, topography, and vegetation-based digital covariates to predict micronutrient content in soil using a stepwise multiple linear regression model in a study area where micronutrient contents showed high variability. They obtained NRMSE values of 87% for Fe and Mn, and 96% for Cu and Zn. Both the reflectance values of the bands and the vegetation indices derived from the ratios of certain bands can reflect the spatial response of vegetation to soil nutrient conditions. Vegetation indices may be sufficient to discriminate between various vegetation characteristics and their associated soil micronutrients concentration. In this respect, the proximity of dates between the time of soil sampling and satellite image acquisition is very important. Digital covariates in Scenarios I and II may be an opportunity for spatial estimation of micronutrient content in rain-fed agricultural fields in arid and semi-arid areas.

It is important to underline the low NRMSE values for all micronutrients produced by the RF model in the scenario using only climate covariates (scenario VIII) (Table 3). This is surprising in the study area where relatively low climatic variation was measured. However, Zeraatpisheh et al. (2020) reported that precipitation that controls soil moisture and water flow, along with digital covariates representing climatic factors, were the most important digital covariates defining the spatial distribution of heavy metals in soils under similar climatic regions in the northeast of Iran. Najafi-Ghiri et al. (2013) reported that the soil moisture regime was an important digital covariate to estimate micronutrient concentration in the calcareous soil of Southern Iran.

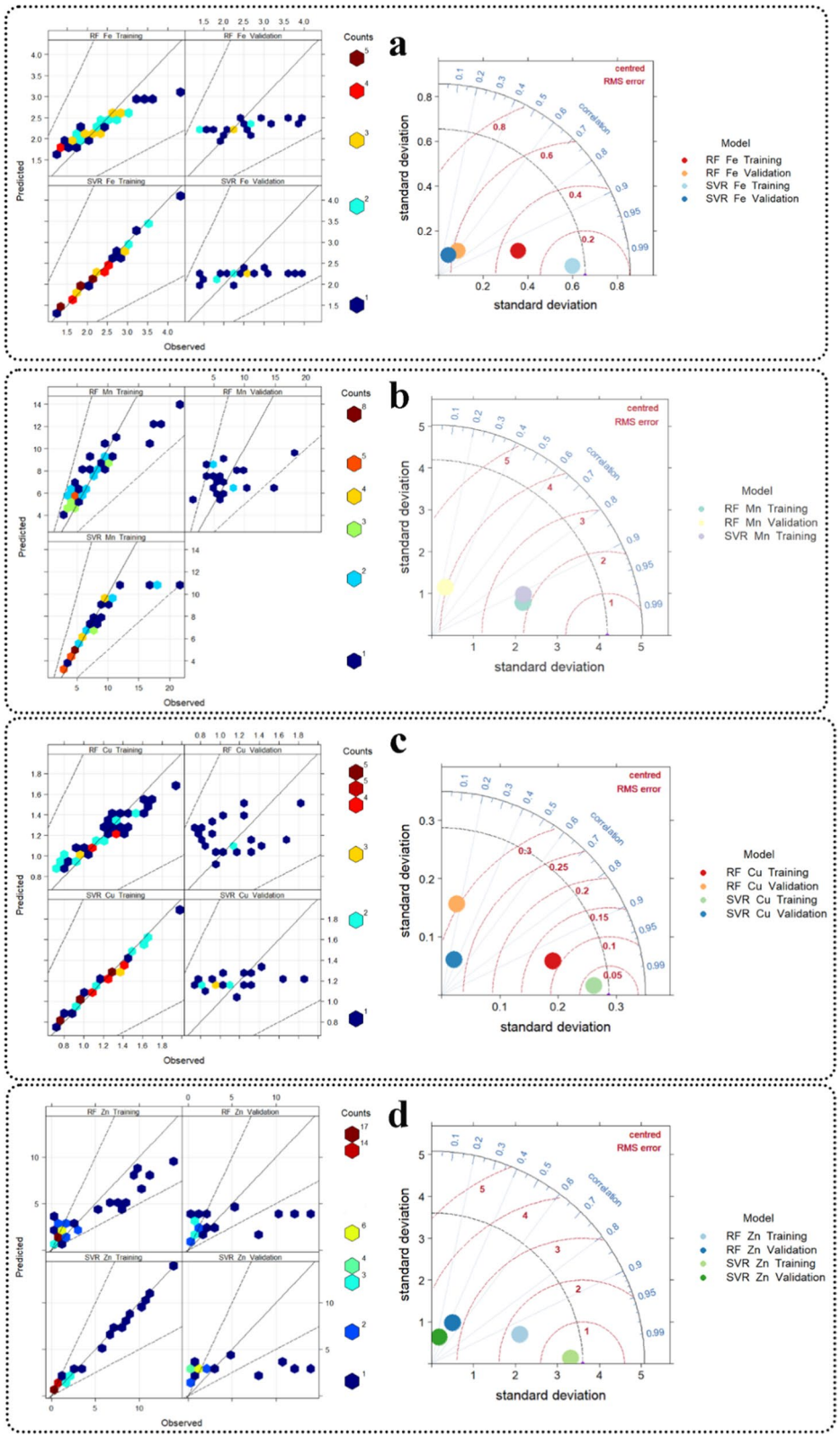
Since measurements of soil properties are expensive and time consuming, it is suggested that scenario III could be applied at the scale of exploratory studies to identify excess hotspots and areas of deficiencies of micronutrients for decision-making. Again, scenario III, which is easily obtained from RS-based freely accessible satellite images, can be used in combination with climate digital covariates (scenario VIII). Like our findings, Azizi et al. (2022) suggested that RS + Topographic + Thematic maps (Geology-Geomorphology) scenario could be used to detect areas that may have heavy metal pollution in exploratory studies in western Iran.

## Controlling factors of micronutrient patterns

While advances have been made in developing effective digital covariates for soil properties of interest (Chen et al. 2022), modelling processes face challenges in a heterogeneous soil environment in which agricultural activities are carried out. Although it was known that soil formation was largely homogeneous at a relatively small spatial scale, and the effects of climate on soil formation were nearly uniform (Zhu et al. 2021), the climate digital covariates of this study were at the forefront of relative importance in the modelling. The low variance climate digital covariates, due to their low spatial resolution (> 100 m), within our study region may have partially masked the effect of other covariates, which may carry information about the local pedological control factors.

Droz et al. (2021) reported that climate digital covariates (i.e., precipitation and temperature) were more important than the other digital covariates such as topographic-based factors in the modelling process. Probably climate digital covariates control soil physicochemical processes, including the redox state of the soil, although their study focused on the estimation of Cu on a European scale.

Climate affects soil properties, vegetation, water retention, soil erosion, and soil organic carbon cycling (Blume et al. 2016a). The soils of the study area are mainly Entisols and Inceptisols, with a low pedogenic age and a strong resemblance to the underlying geological material. The mineral structure of the rocks and parent materials is shaped by the climate and contains micronutrients within the constraints of this structure. Differences in climatic factors that directly affect soil moisture can have significant effects on the valent forms of micronutrients in local areas (Blume et al. 2016b; Droz et al. 2021). Precipitation can temporarily increase the soil water content and reduce the oxygen diffusion in the water, leading to a decrease in available Cu. Under anaerobic conditions, biological activity can reduce Cu mobility by precipitating soluble Cu (II) to Cu (I). The temporal availability or absence of water can affect the Mn levels (Denton-Thompson and Sayer 2022). The BIO-15- Precipitation Seasonality was determined as an important digital covariate during the modelling process with the SVR and RF algorithms for Mn. The concentration of Mn largely depends


a
RF Fe Training
RF Fe Validation
SVR Fe Training
SVR Fe Validation
Predicted
Observed
Counts
centred
RMS error
correlation
standard deviation
Model
RF Fe Training
RF Fe Validation
SVR Fe Training
SVR Fe Validation
b
RF Mn Training
RF Mn Validation
SVR Mn Training
RF Mn Training
RF Mn Validation
SVR Mn Training
c
RF Cu Training
RF Cu Validation
SVR Cu Training
SVR Cu Validation
d
RF Zn Training
RF Zn Validation
SVR Zn Training
SVR Zn Validation

◂**Fig. 3** Model results obtained using Scenario XI. **a** Iron (Fe), **b** Manganese (Mn), **c** Copper (Cu), **d** Zinc (Zn). In the Taylor diagram are reported the measured standard deviation (purple point), the contour of constant root mean squared error (RMSE, red circular dashed arcs), the contour of constant standard deviation (black circular dashed arc), and the contour of constant correlation coefficient (blue line). In the scatter plots, "counts" is the sample number

on the soil pH, redox conditions, and active Mn reserves in the soil (Blume et al. 2016a). The most important process affecting redox conditions may be the periodicity of precipitation which was represented by BIO-15- Precipitation Seasonality digital covariate.

The effects of topographic factors location, slope, and elevation on soil micronutrient variability have been outlined in Adhikari et al. (2018). Under a Piedmont plain, topographic position can poorly affect the spatial distributions of soil micronutrients in different land uses. Specifically, areas with higher topographic wetness index (TWI) values tend to receive more overland runoff and may exhibit higher moisture levels compared to the surrounding region. This can lead to the deposition of micronutrients through runoff processes. However, topography had a relatively smaller effect on the prediction of the micronutrients because the amount of movement and leaching of the micronutrients was very poor, especially under slight to moderate alkaline conditions in the study area as shown in Table 2.

Despite the same of the parent material, deposition and transport of soil, and different weather conditions changing at the micro-scale, management practices in different land use may affect micronutrient availability and distribution.

### Challenges and future perspectives for mapping soil micronutrients

Much work has been done on modelling studies to understand the processes that regulate nutrient availability, soil properties, and environmental conditions (Heng et al. 2017; FAO 2022a). Accordingly, soil mapping efforts should be based on a soil monitoring system that can accurately represent real soil conditions over time and at an appropriate scale. In the progress of data-driven soil mapping, the fact that the soil surface is covered during most of the year can be seen as a limitation, especially in agricultural production areas (Tziolas et al. 2021). If the soil surface is covered, the integration of open-access digital data into the DSM process should be organised in such a way that it can spatially and radiometrically explain the heterogeneity in nutrient levels of crops grown on the land. Sufficient heterogeneity of vegetation characteristics captured by satellite imagery-based vegetation indices may contain more information on soil micronutrient content, since vegetation cover is often a comprehensive expression of various factors such as soil and terrain. Higher spatial resolution may have a more critical value in estimating the spatial variability of soil properties.

Future global initiatives (FAO 2022b) were recommended to be highly sensitive in terms of linking to decision-making and to include information on how to use it for end-user farmers (Chagumaira et al. 2022).

## Conclusions

This study characterised spatial patterns in micronutrient content and the accompanying modelling uncertainties for part of the piedmont plain of northern Iran. Although the spatial variation of the climate digital covariates was low, it produced usefulness results for estimating the micronutrient content. The main findings of the spatial modelling of micronutrient contents in an agricultural system in an arid region were as follows.

In the final scenario (scenario XI), which was prepared with the assistance of expert opinion, the Random Forests model proved to be better than the Support Vector Regression model as it was able to reproduce the heterogeneity of the micronutrients and the extreme values using the scenario XI produced maps.

Integrated scenarios using Landsat 8 OLI and Sentinel-2 time series, climate digital covariates, and machine learning algorithms (scenario XI) can be progressed by applying them for exploratory studies in similar study areas where no soil micronutrient information is currently available.

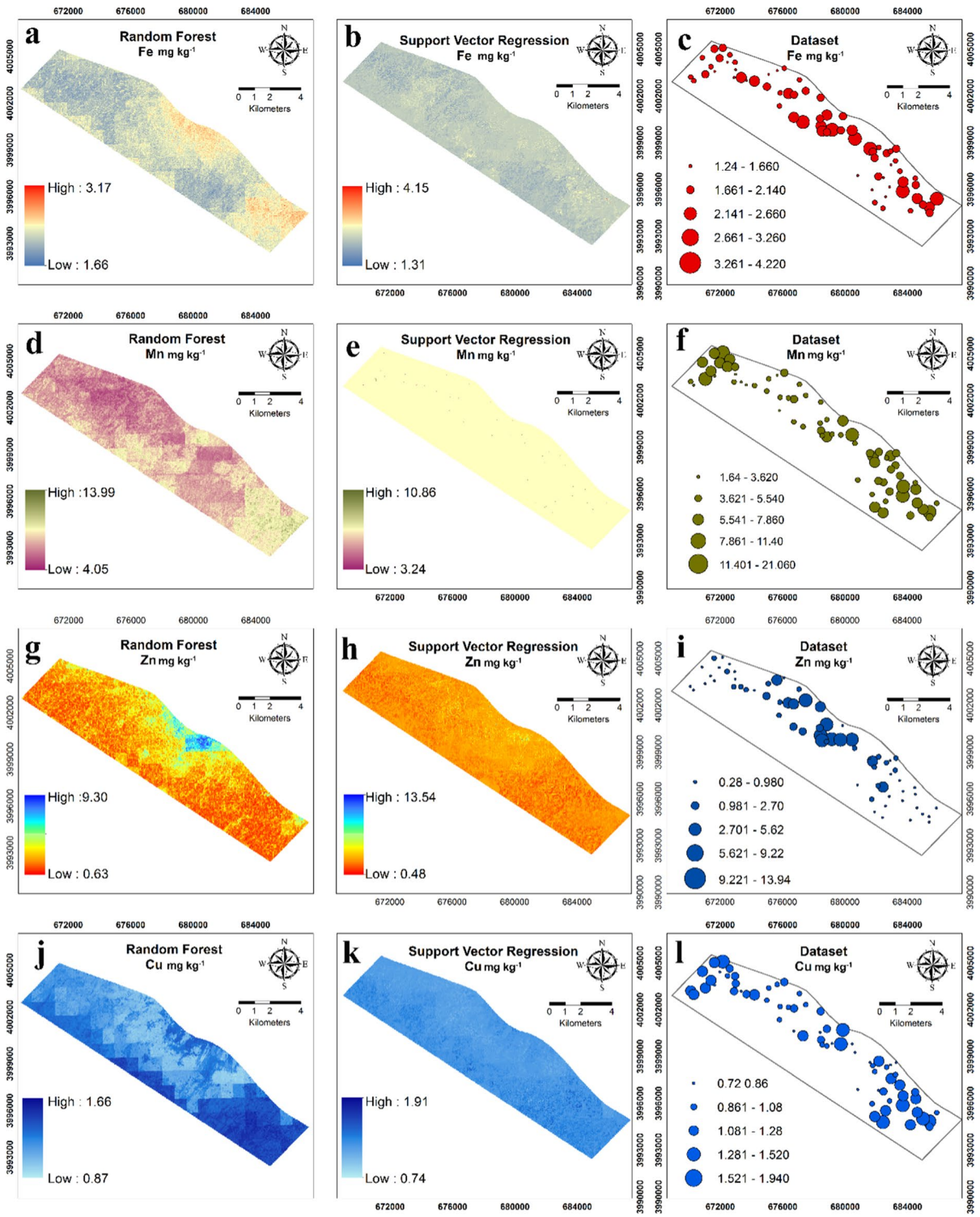



**Fig. 4** Spatial distribution of soil micronutrients predicted by ML models using Scenario XI and measured in the reference dataset: Fe using RF model (**a**); Fe using SVR model (**b**); Fe measured (**c**); Mn using RF model (**d**); Mn using SVR model (**e**); Mn measured (**f**); Zn using RF model (**g**); Zn using SVR model (**h**); Zn measured (**i**); Cu using RF model (**j**); Cu using SVR model (**k**); Cu measured (**l**)

In the future, the presence of soil scientists will be important in the process from the identification to the selection of specific digital covariate sets for micronutrients.

**Acknowledgements** The authors gratefully acknowledge the financial support provided by the Department of Soil Science, University of Tehran, Iran (research project no. 4886791).

**Author contributions** Resources, conceptualisation writing-reviewing, and editing: AK. Conceptualisation, data curation, formal analysis, methodology, visualisation, writing-original draft preparation, writing-reviewing, and editing: FK. Visualisation, writing-reviewing and editing: LB, YG-A, JRC, AC-C.

**Data availability statement** The datasets generated during and/or analysed during the current study were available in the [Scenario_based_DSM_Micronutrients] repository, [https://doi.org/10.5281/zenodo.7008079].

**Declarations**

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Adhikari K, Owens PR, Ashworth AJ, Sauer TJ, Libohova Z, Richter JL, Miller DM (2018) Topographic controls on soil nutrient variations in a silvopasture system. Agrosyst Geosci Environ 1:1–15. https://doi.org/10.2134/age2018.04.0008

Alloway BJ (2013) Heavy metals and metalloids as micronutrients for plants and animals. In: Alloway B (eds) Heavy Metals in Soils. Environmental Pollution, vol 22. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4470-7_7

ALOS PALSAR (2021) Dataset:© JAXA/METI ALOS PALSAR L1.0 2007. Accessed through ASF DAAC 05 September 2021

Azizi K, Ayoubi S, Nabiollahi K, Garosi Y, Gislum R (2022) Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. J Geochem Explor 233:106921. https://doi.org/10.1016/j.gexplo.2021.106921

Bagherzadeh A, Ghadiri E, Souhani Darban AR, Gholizadeh A (2016) Land suitability modeling by parametric-based neural networks and fuzzy methods for soybean production in a semi-arid region. Model Earth Syst Environ 2:1–11. https://doi.org/10.1007/s40808-016-0152-4

Blume HP, Brümmer GW, Fleige H, Horn R, Kandeler E, Kögel-Knabner I, Wilke BM (2016a) Soil-plant relations. In: Blume HP, Brümmer GW, Fleige H, Horn R, Kandeler E, Kögel-Knabner I, Wilke BM (eds) Scheffer/SchachtschabelSoil Science. Springer, Berlin, Heidelberg, pp 409–484 https://doi.org/10.1007/978-3-642-30942-7_9

Blume HP, Brümmer GW, Fleige H, Horn R, Kandeler E, Kögel-Knabner I, Wilke BM (2016b) Soil Organic Matter. In: Blume HP, Brümmer GW, Fleige H, Horn R, Kandeler E, Kögel-Knabner I, Wilke BM (eds) Scheffer/Schachtschabel soil science, Springer, Berlin, Heidelberg, pp 55–86 https://doi.org/10.1007/978-3-642-30942-7_3

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Brown KS, Libohova Z, Boettinger J (2017) Digital Soil Mapping. In: Ditzler C, Scheffe K, Monger HC (eds) Soil survey manual, USDA Handbook 18. Government Printing Office, Washington, D.C, pp 295–354

Carr D, Lewin-Koh N, Maechler M, Sarkar D (2021) hexbin: Hexagonal Binning Routines. https://CRAN.R-project.org/package=hexbin

Carslaw DC, Ropkins K (2012) openair-an R package for air quality data analysis. Environ Model Softw 27–28:52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

Chagumaira C, Chimungu JG, Nalivata PC, Broadley MR, Nussbaum M, Milne AE, Lark RM (2022) Mapping soil micronutrient concentration at national-scale: an illustration of a decision process framework. Egusphere. https://doi.org/10.5194/egusphere-2022-583

Chen S, Arrouays D, Mulder VL, Poggio L, Minasny B, Roudier P, Walter C (2022) Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. Geoderma 409:115567. https://doi.org/10.1016/j.geoderma.2021.115567

Cortes C, Vapnik V (1995) Support-Vector Networks. Mach Learn 20:273–297. https://doi.org/10.1007/BF00994018

Denton-Thompson SM, Sayer EJ (2022) Micronutrients in food production: what can we learn from natural ecosystems? Soil Syst 6:8. https://doi.org/10.3390/soilsystems6010008

Droz B, Payraudeau S, Rodríguez Martín JA, Tóth G, Panagos P, Montanarella L, Borrelli P, Imfeld G (2021) Copper content and export in european vineyard soils influenced by climate and soil properties. Environ Sci Technol 55:7327–7334. https://doi.org/10.1021/acs.est.0c02093

Drucker H, Burges CJC, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. Adv Neur in 9:155–161

ESA (2015) European Space Agency. Sentinel-2 User Handbook Rev 2. Available from: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000

ESRI (2021) ArcGIS user's guide, http://www.esri.com

FAO (2022a) Soils for nutrition: state of the art. Rome. 78 pp. Accessed on link: https://doi.org/10.4060/cc0900en. Accessed date: 18/08/2022a

FAO (2022b) FAO's Global Soil Partnership kicks start a country-driven process to map soil nutrients and soil nutrient budgets at all scales. Accessed on link: https://www.fao.org/global-soil-partnership/resources/highlights/detail/en/c/1601502/

FAO and ITPS (2018) Global Soil Organic Carbon Map (GSOCmap) Technical Report. Rome. 162 pp. Accessed on link: https://www.fao.org/3/I8891EN/i8891en.pdf

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. Int J Climatol 37:4302–4315. https://doi.org/10.1002/joc.5086

Foroughifar H, Jafarzadeh AA, Torabi H, Pakpour A, Miransari M (2013) Using geostatistics and geographic information system techniques to characterize spatial variability of soil properties, including micronutrients. Commun Soil Sci Plant Anal 44:1273–1281. https://doi.org/10.1080/00103624.2012.758279

Gitelson AA, Kaufman YJ, Merzlyak MN (1996) Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens Environ 58:289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gopp NV, Savenkov OA (2019) Relationships between the NDVI, yield of spring wheat, and properties of the plow horizon of eluviated clay-illuvial chernozems and dark gray soils. Eurasian Soil Sci 52:339–347. https://doi.org/10.1134/S1064229319030050

Gruber S, Peckham S (2009) Land-surface parameters and objects in hydrology. In: Heng T, Reuter HI (eds), Developments in soil science, vol 33, pp 171–194 https://doi.org/10.1016/S0166-2481(08)00007-X

Hengl T, Leenaars JG, Shepherd KD, Walsh MG, Heuvelink G, Mamo T, Kwabena NA (2017) Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutr Cycl Agroecosyst 109:77–102. https://doi.org/10.1007/s10705-017-9870-x

Hengl T, Reuter HI (2008) Geomorphometry: concepts, software, applications. In: Soil science. vol. 33. Elsevier, Amsterdam

Hijmans RJ (2020) Raster: Geographic Data Analysis and Modeling. R package version 3.4–5. https://CRAN.R-project.org/package=raster

Hounkpatin KO, Schmidt K, Stumpf F, Forkuor G, Behrens T, Scholten T, Welp G (2018) Predicting reference soil groups using legacy data: a data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). Sci Rep 8:1–16. https://doi.org/10.1038/s41598-018-28244-w

Kaya F (2022) fuatkaya/Scenario_based_DSM_Micronutrients: Scenario_based_DSM_Micronutrients_Zenodo https://doi.org/10.5281/zenodo.7008080 (Released on 2022-08-18)

Kaya F, Başayiğit L (2022) Using machine learning algorithms to mapping of the soil macronutrient elements variability with digital environmental data in an alluvial plain. In: Artificial intelligence and smart agriculture applications, Auerbach Publications, pp 107–136 https://doi.org/10.1201/9781003311782-6

Keshavarzi A, Kaya F, Kaplan G, Başayiğit L (2022) Land cover classification in an arid landscape of Iran using Landsat 8 OLI science products: Performance assessment of machine learning algorithms. 4thIntercontinental Geoinformation Days (IGD), pp 175–179. Tabriz, Iran. https://publish.mersin.edu.tr/index.php/igd/article/view/586

Kuhn M (2020) caret: Classification and Regression Training. R package version 6.0–86. https://CRAN.R-project.org/package=caret

Lark RM, Ander EL, Cave MR, Knights KV, Glennon MM, Scanlon RP (2014) Mapping trace element deficiency by cokriging from regional geochemical soil data: a case study on cobalt for grazing sheep in Ireland. Geoderma 226:64–78. https://doi.org/10.1016/j.geoderma.2014.03.002

Liang P, Qin C, Zhu A, Hou Z, Fan N, Wang Y (2020) A case-based method of selecting covariates for digital soil mapping. J Integr Agric 19:2127–2136. https://doi.org/10.1016/S2095-3119(19)62857-1

Liang P, Qin CZ, Zhu AX (2021) Comparison on two case-based reasoning strategies of automatically selecting terrain covariates for digital soil mapping. Trans GIS 25:2419–2437. https://doi.org/10.1111/tgis.12831

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2:18–22

Lindsay WL, Norvell W (1978) Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Sci Soc Am J 42:421–428. https://doi.org/10.2136/sssaj1978.03615995004200030009x

Malone B, Arrouays D, Poggio L, Minasny B, McBratney A (2022) Digital soil mapping: evolution, current state and future directions of the science. Ref Module Earth Syst Environ Sci. https://doi.org/10.1016/B978-0-12-822974-3.00130-0

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2020) e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7–4

Miran N, Rasouli Sadaghiani M, Feiziasl V, Sepehr E, Rahmati M, Mirzaee S (2021) Predicting soil nutrient contents using Landsat OLI satellite images in rain-fed agricultural lands, northwest of Iran. Environ Monit Assess 193:1–12.

Mponela P, Snapp S, Villamor G, Tamene L, Le QB, Borgemeister C (2020) Digital soil mapping of nitrogen, phosphorus, potassium, organic carbon and their crop response thresholds in smallholder managed escarpments of Malawi. Appl Geogr 124:102299. https://doi.org/10.1016/j.apgeog.2020.102299

Naimi S, Ayoubi S, Demattê JAM, Zeraatpisheh M, Amorim MT, Mello FAO (2021) Spatial prediction of soil surface properties in an arid region using synthetic soil image and machine learning. Geocarto Int. https://doi.org/10.1080/10106049.2021.1996639

Najafi-Ghiri M, Ghasemi-Fasaei R, Farrokhnejad E (2013) Factors affecting micronutrient availability in calcareous soils of Southern Iran. Arid Land Res Manag 27:203–215.

Nussbaum M (2022) Machine learning and processing of large data. Ref Module Earth Syst Environ Sci.

Pasolli L, Notarnicola C, Bruzzone, (2011) Estimating soil moisture with the support vector regression technique. IEEE Geosci Remote Sens 8:1080–1084. https://doi.org/10.1109/LGRS.2011.2156759

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Sayler K, Zanter K (2021) Landsat 8 Collection 2 (C2) Level 2 Science Product (L2SP) Guide LSDS-1619 Version 2.0, EROS Sioux Falls, South Dakota

Sharma RP, Chattaraj S, Jangir A, Tiwari G, Dash B, Daripa A, Naitam RK (2022) Geospatial variability mapping of soil nutrients for site specific input optimization in a part

of Central India. Agronomy J 114:1489–1499. https://doi.org/10.1002/agj2.21025
Smith WK, Dannenberg MP, Yan D, Herrmann S, Barnes ML, Barron-Gafford GA, Yang J (2019) Remote sensing of dryland ecosystem structure and function: progress, challenges, and opportunities. Remote Sens Environ 233:111401. https://doi.org/10.1016/j.rse.2019.111401
Snapp S (2022) Embracing variability in soils on smallholder farms: New tools and better science. Agric Syst 195:103310. https://doi.org/10.1016/j.agsy.2021.103310
Soil Survey Staff (2014) Keys to soil taxonomy, 12th edn. USDA-Natural Resources Conservation Service, Washington, DC
Srisomkiew S, Kawahigashi M, Limtong P, Yuttum O (2022) Digital soil assessment of soil fertility for Thai jasmine rice in the Thung Kula Ronghai region, Thailand. Geoderma 409:115597. https://doi.org/10.1016/j.geoderma.2021.115597
Tziolas N, Tsakiridis N, Chabrillat S, Demattê JAM, Ben-Dor E, Gholizadeh A, Zalidis G, Wesemeal B (2021) Earth observation data-driven cropland soil monitoring: a review. Remote Sens 13:4439. https://doi.org/10.3390/rs13214439
Vasu D, Sahu N, Tiwary P, Chandran P (2021) Modelling the spatial variability of soil micronutrients for site specific nutrient management in a semi-arid tropical environment. Model Earth Syst Environ 7:1797–1812. https://doi.org/10.1007/s40808-020-00909-4
Wadoux AMC, Minasny B, McBratney AB (2020) Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci Rev 210:103359. https://doi.org/10.1016/j.earscirev.2020.103359
Wadoux AMC, Odeh IOA, McBratney AB (2021) Overview of Pedometrics. Ref Module Earth Syst Environ Sci. https://doi.org/10.1016/B978-0-12-822974-3.00001-X
Wadoux AMC, Walvoort DJJ, Brus DJ (2022) An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams. Geoderma 405:115332. https://doi.org/10.1016/j.geoderma.2021.115332
Wang SY, Yu TQ, Wang JL, Liu YANG, Kai YANG, Ping LU (2008) Preliminary study on spatial variability and distribution of soil available microelements in Pinggu County, Beijing, China. Agr Sci in China 7:1235–1244. https://doi.org/10.1016/S1671-2927(08)60170-4
Wilding L (1985) Spatial variability: its documentation, accommodation and implication to soil surveys. In: Soil spatial variability. Workshop, pp 166–194.
Xiao J, Shen Y, Tateishi R, Bayaer W (2006) Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. Int J Remote Sens 27:2411–2422. https://doi.org/10.1080/01431160600554363
Yigini Y, Olmedo GF, Reiter S, Baritz R, Viatkin K, Vargas R (2018) Soil Organic Carbon Mapping Cookbook 2nd edition. Rome, FAO. 220 pp. Available from: https://www.fao.org/3/I8895EN/i8895en.pdf
Zambrano-Bigiarini M (2020) hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series R package version 0.4-0. URL https://github.com/hzambran/hydroGOF. https://doi.org/10.5281/zenodo.839854
Zeraatpisheh M, Mirzaei R, Garosi Y, Xu M, Heuvelink GBM, Scholten T, Taghizadeh-Mehrjardi R (2020) Feasibility of using environmental covariates and machine learning to predict the spatial variability of selected heavy metals in soils, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-12378. https://doi.org/10.5194/egusphere-egu2020-12378
Zhang X, Li MJ, Zhan LQ, Zhan LQ, Wu W, Liu HB (2020) Boron availability in top- and sub-soils as affected by topography and climate. Nutr Cycl Agroecosyst 118:91–101. https://doi.org/10.1007/s10705-020-10085-7
Zhu H, Ding H, Bi R, Hou M (2021) Large-extent spatial heterogeneity of soil bioavailable micronutrients and the relative roles of environmental indicators on them within maize fields. Ecol Indic 130:108071. https://doi.org/10.1016/j.ecolind.2021.108071