Check for updates

Soil moisture data using citizen science technology cross-validated by satellite data

Mohammad Karamouz ⓘ *, Elham Ebrahimi and Arash Ghomlaghi

School of Civil Engineering, University of Tehran, Tehran, Iran
*Corresponding author. E-mail: karamouz@ut.ac.ir

ⓘ MK, 0000-0002-6573-262X

ABSTRACT

Soil moisture represents many attributes of the geo-hydrological cycle and the climate system. Citizen science through social media as an emerging tool could be utilized to collect soil moisture data. A pilot study area was selected in Shahriar, Iran. A user interface and a sampling process (use of citizen science by subscribers) were designed to analyze the subjective and gravimetric soil moisture data. Furthermore, explanatory moisture condition (EMC), a new initiative to consider land use in soil moisture information from vegetation cover, was evaluated. A statistical artificial neural network was used for quantifying subjective data, and soil moisture layouts were produced by utilizing the ordinary kriging (OK) method. For cross-validating, the land surface temperature data from the MODIS satellite were retrieved. A platform for the region with 200 m grids resolution to collect daily soil moisture at eight ungauged stations is proposed to utilize subjective data from the subscribers and cross-validated with satellite data. A virtual station at the centroid of the pervious part of the study area was selected as a reference station for data collection daily or weekly to generate soil moisture time series. The results showed a high potential of utilizing satellite and citizen science data for real-time estimation of scarce soil moisture data in developing regions.

Key words: citizen science, data-driven modeling, image processing, kriging method, soil moisture

HIGHLIGHTS

- Design of a platform for real-time data estimation of SM in virtual station(s) using color contrast image processing with simple user interface for retrieving SM data through social media.
- Cross-validation and error analysis with daily downscaled satellite SM data provides a unique opportunity for SM estimation in the developing regions with no national or regional plan to collect time series of this data.

1. INTRODUCTION

Soil moisture is one of the essential components of climate, ecological, and the hydrological analysis and modeling (MEA 2005) and also is the best indicator of climate change (Karamouz et al. 2019). It plays a vital role in exchanging water, energy, and carbon between soil and air, even though the unsaturated soil's water content is less than 0.15% of the total global available freshwater (Dobriyal et al. 2012). Moreover, soil moisture is a crucial factor in the formation of floods and droughts. It can be used for predicting and monitoring these events (García et al. 2019). Furthermore, soil moisture is commonly analyzed to increase irrigation efficiency and prevent dust storms (Ziadat & Taimeh 2013; Kim & Choi 2015).

As one of the most critical hydrological variables, soil moisture, contrary to variables such as temperature, precipitation, and streamflow, does not have a custodian in many developing regions (Dripps & Bradbury 2007; Wu et al. 2021). Soil moisture can be measured/estimated by different methods, such as field measurements, remote sensing, and atmospheric–terrestrial water-balance estimations (ATWB). Conventional in-situ soil moisture measurements are time-consuming, costly, and require maintenance and repair (Rinderer et al. 2015). Satellite soil moisture data have more potential for large-scale modeling. Although these data have high temporal variability, their spatial resolution is low. Therefore, they need to be downscaled and verified by field measurements (Wang & Qu 2009). However, there are some satellites, as well as Sentinel-1, which are designed to work in a pre-programmed operation mode for imaging high-resolution coastal zones, and polar areas (Torres et al. 2012). Wilson et al. (2003) pointed out the importance of soil moisture content in the root zone. They tried to establish a relationship between soil moisture content over two depths of 0–6 and 0–30 centimeters.

This relationship's significance is that the remotely sensed data represent the moisture content at a few centimeters below the soil. Dobriyal *et al.* (2012) investigated available methods for estimating soil moisture and its implications for water resource management. They compared existing methods based on several indicators, such as cost-effectiveness, accuracy, spatial scale, and response time.

One of the common characteristics of soil moisture is its spatial and temporal variability. The traditional solution to address this problem is to make various field measurements over the catchment at different times, which is very costly and human resources intensive. Thus, utilizing a citizen science-based framework can help to solve this problem. Citizen science is one of the social media tools that could be utilized to collect soil moisture data. The essential element of citizen science is the participation of non-scientists in scientific research. Typically, citizens act as observers or experimenters within the structures established by professional scientists. Using citizen science to estimate soil moisture involves citizens observing soil and landscape by increasing the density and geographic spread of observations, increasing the accuracy of mapping the data (Rossiter *et al.* 2015). Rinderer *et al.* (2012) developed a citizen science-based method for classifying soil wetness in wet environments. After designing a questionnaire, they gathered qualitative data and evaluated the effectiveness of their proposed method. Their focus was on assessing the consensus of users on the choice of wetness class. Although they compared the questionnaires' data with the field measurements (soil moisture data derived by the gravimetric and time domain reflectometry method), they did not quantify and use the gathered data. Njue *et al.* (2019) assessed citizen science in hydrological monitoring and ecosystem services management. They provided a comprehensive review of citizen science and crowd sourced data collection focusing on the role of citizen science in generating a plethora of scientific data.

Khadim *et al.* (2020) used citizen science to complement in-situ data scarcity in the context of groundwater management in the Gilgel-Abay region, Ethiopia. They consider the collections of locally trained high school students and farmers to be verified by faculty experts.

Several software tools have been developed for using citizen science in order to monitor the environment. Crowd water, which is developed at the department of geography, the University of Zurich (www.crowdwater.ch), is a mobile application through which users can record data related to soil moisture, water level, and temporary streams. App EAR is another example of these software tools (www.app-ear.com.ar). This application, which is designed at the National University of La Plata, aims to collect data related to water bodies such as water transparency and smell, aquatic plants, and riverside vegetation. Flood crowd, which is developed at Loughborough University, is a web-based data collection system to assess the flood risk and damage (www.citizensciencecenter.com).

Analyzing the information gathered by citizen scientists requires knowledge about soil characteristics. Different soils have different compositions and physical properties. Meanwhile, the soil's color can be a comprehensive indicator of the soil's physical characteristics and chemical compositions, such as soil moisture, mineral composition, and organic content. Thus, a significant portion of soil properties can be obtained from soil color (Han *et al.* 2016). The darkness of the soil is a characteristic that depends on the amount of soil moisture. Therefore, image processing could be used to analyze the soil color and estimate soil moisture. This technique was utilized in this study. It performs some operations on an image to produce an enhanced image or to retrieve some useful information. It is a signal/image processing type in which input is a digital image, and output may be associated with that image. Nowadays, image processing is among the rapidly growing technologies. Over the past decades, opportunities have become available for analyzing measurements using image processing techniques to develop high-speed and resolution computers. Estimating soil moisture using image processing is an area that has a high potential for exploring possibilities for rapid, accurate, and non-destructive quantification. Heil *et al.* (2019) utilized a k-means clustering algorithm to classify diffuse reflectance soil spectra in West African soils. In their study, over 1,000 mid-infrared diffuse reflectance Fourier transform spectra of agricultural soils were collected. The clustering process was done by the k-means method to explore the feasibility of centroid-based clustering algorithms to find the spectral data.

Antecedent moisture condition (AMC) is a significant factor determining the abstraction estimations in the SCS-CN method. The AMC classification is done in three levels: dry (AMC I), average (AMC II), and wet (AMC III) based on the five-day antecedent rainfall amount both in dormant and growing seasons: see Karamouz *et al.* (2012) for more details. In the semi-arid regions, rainfalls are sporadic and have high spatial variations, so soil moisture estimation may not be practical in this study. However, explanatory data offers some merits if subjective information can play an important role in citizen science exploration of soil moisture applications. Karamouz & Fereshtehpour (2019) used explanatory data for nonstationary spatial data for flood inundation in NYC. In this study, a simple distinction

of Explanatory Moisture Condition is made based on land use to separate the vacant/open space or developed/residential (EMC 1) from vegetated/agricultural areas (EMC 2). This principle is similar to the Antecedent Moisture Condition (AMC) attribute of SCS (Soil Conservation Service). However, both terms have different connotations and implications. The intention is not to undermine the well-established AMC stature in the rainfall-runoff analysis. In this study, the EMC initiative could open many local and subjective elements in an environment for further study of citizen science.

EMC is utilized in conjunction with ANN input to represent the relationship between dependent variables and auxiliary data (land use). Remote sensing methods offer the best means of soil moisture estimation at a large spatial scale. At present, soil moisture can be retrieved by a variety of remote sensing techniques. Microwave technology has demonstrated a quantitative ability to estimate soil moisture physically for a wide range of vegetation cover. The Soil Moisture and Ocean Salinity (SMOS) (Kerr *et al.* 2001) and the Soil Moisture Active Passive (SMAP) (Entekhabi *et al.* 2010) are two of the most recent satellite missions for global soil moisture retrieval using microwave remote sensing. However, current satellite microwave radiometers' spatial resolutions are not suitable for land remote sensing, especially surface soil moisture (SSM) monitoring. This is due to practical problems related to supporting a large, low-frequency antenna in space. Methods with optical, thermal infrared remote sensing (TIR) measurements from Terra/Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) and Land Remote Sensing (Landsat) have been used for SSM monitoring. Also, Thematic Mapper (TM)/ Enhanced Thematic Mapper (ETM)/Thermal Infrared Sensor (TIRS) data are other techniques available at moderate or high spatial resolutions. These are based on the universal triangular relationship between land surface temperature (LST) and the normalized difference vegetation index (NDVI) (Sandholt *et al.* 2002; Sun *et al.* 2012). Xu *et al.* (2018) used remote sensing measurements to estimate surface soil moisture at both high spatial and temporal resolutions by using Moderate Resolution Imaging Spectroradiometer (MODIS) and Land Remote Sensing Satellite (Landsat) data in the state of Iowa. They developed a nonlinear regression model to retrieve the surface soil moisture. They used the land surface temperature (LST) and normalized difference vegetation index (NDVI) data from May 1, 2016, to August 31, 2016. Detecting the spatial variations of soil moisture is a crucial step towards improving the accuracy of hydrological models. Besides, soil moisture maps are of paramount importance in the spatiotemporal monitoring of this variable. There are several geostatistical methods to model soil properties' spatial variability (Zare-Mehrjardi *et al.* 2010). Geostatistical methods, such as simple kriging and ordinary kriging (OK), utilized in this study, could be used to develop soil moisture maps from discrete/point measurements of this variable. While simple kriging assumes a known mean over the entire domain, OK assumes a constant unknown mean only over the search neighborhood. Sulieman & Algarni (2019) used the OK technique to predict and map soil organic carbon at several depth intervals in alluvium soils along the Blue Nile, Sudan. They collected 152 samples at several depths and utilized a spherical semivariogram to describe the data's autocorrelation. The results revealed that the OK model was correctly estimating the variability of soil organic carbon in the study area. Friedland *et al.* (2017) used isotropic and anisotropic kriging approaches for estimating surface-level wind speeds resulting from windstorms across large and geographically diverse regions. The results revealed that anisotropic kriging is appropriate to use for interpolating wind speeds because it accounts for wind direction and trends in wind speeds.

Soil moisture is one of the most critical hydrological variables. It is challenging in many developing regions as it does not have a custodian similar to rainfall and river flow data.

To the best of our knowledge and based on a rather extensive literature search, the application of citizen science for soil moisture data collection and cross-validated by satellite data has not been made. The application of multi-directional kriging (MDK) and the introduction of explanatory soil moisture condition and a platform is proposed in this paper for collecting subjective soil moisture data using citizen science subscribers. This is done at selected ungauged locations with the capability for regionalizing data at 200 m grid resolution.

The methodological and theoretical approach in this study, is to combine subjective information collected in the field with the satellite data to develop a platform for SM estimation utilizing explanatory data and image processing. The subjective information is collected from citizen science (social media) and the satellite data is from MODIS land surface temperature product to estimate soil moisture data. In the absence of an official soil moisture measuring station, the proposed methodology could serve as an effective tool to collect this much-needed hydrogeological data. In the Materials and Methods section, different aspects of the proposed methodology are described. Then, results are presented, and finally, conclusions are given.

2. MATERIALS AND METHODS

This research consists of four main steps to design and carry out this study as shown in Figure 1. It should be noted that the analysis part of this study consists of spatial, image processing, data driven (ANN), and soil moisture estimation analysis that are also depicted in this figure. In the first step, the user interface was developed. Then in the second step, the first sampling process was organized. Gravimetric, citizen science, explanatory moisture condition (EMC), and satellite land surface temperature data were collected. EMC classifications are EMC I for non-vegetated and EMC II for vegetated areas. In the third step, the collected data was analyzed, and two ANN models were developed to quantify collected subjective data. Based on the correlation between soil moisture and land surface temperature, an expression was developed to estimate soil moisture based on MODIS LST data. Then, soil moisture layouts were produced by the OK method, and by performing a second sampling, cross-validation of SM data between citizen science data, and satellite data, was conducted. Finally, a soil moisture estimation platform was developed to estimate SM in the scarce developing region. The last box in this study is SM product in the form of a single SM estimate in a given day at the virtual station or a soil moisture layout for the study area. It should be noted that the necessary hardware, software, and organizational setup should be in place before claiming that this is for a real-time application.

2.1. Description of the user interface

The soil moisture subjective assessment form was developed in a local application site (App.) called Porsline (https://porsline.ir). This online application has features such as making a form and questionnaire with a few drags and drop, ready-made
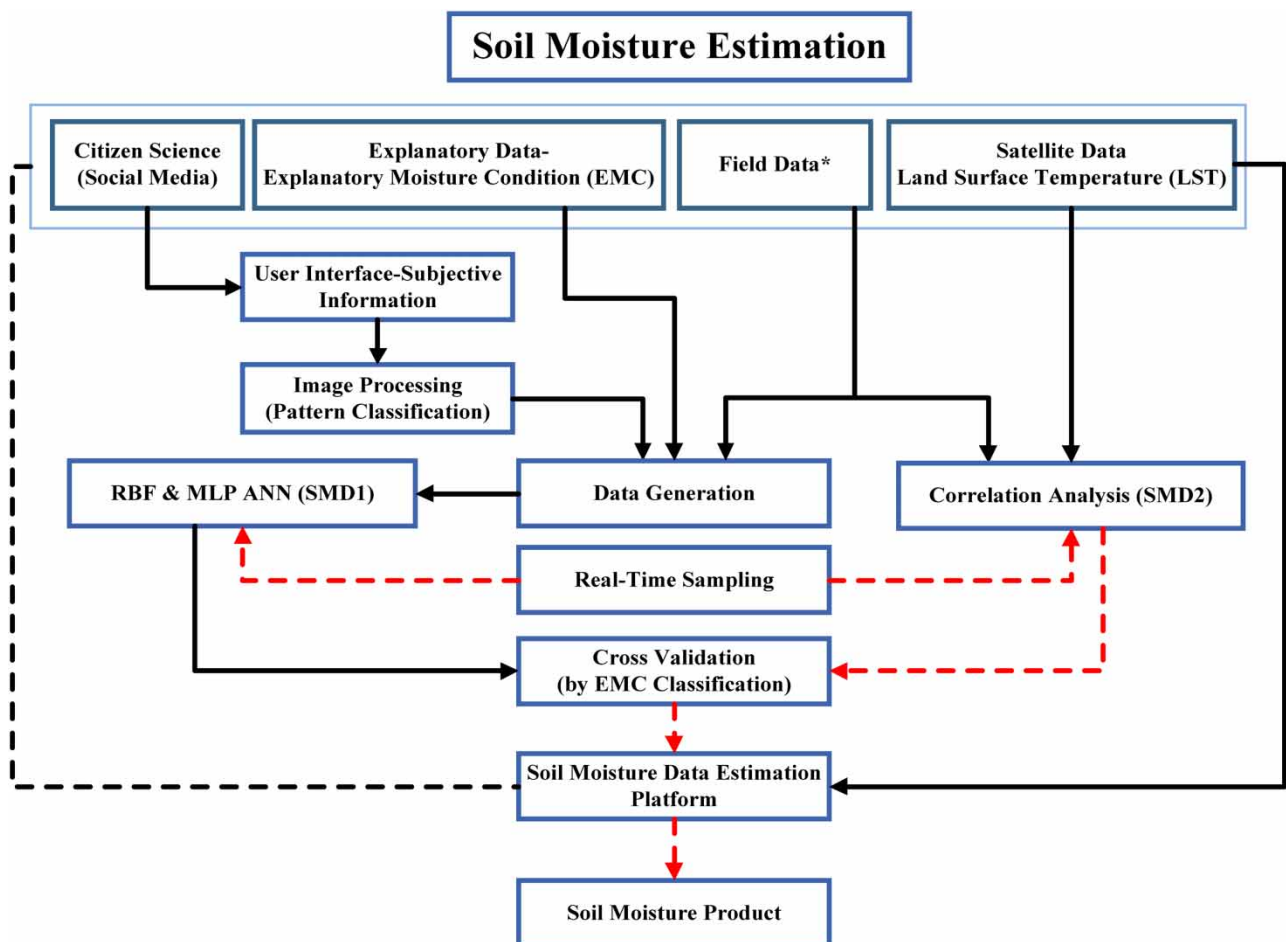
*Field Data was not collected during real-time sampling

Figure 1 | The proposed methodology for soil moisture estimation with real-time dash line attributes.

questionnaires, intelligent display of questions, and forms compatible with all types of mobile phones and tablets. It was used to create the user interface for this study. After posting their personal information and sampling specifications (spatial coordinates and sampling time), users can respond to the soil moisture questions. They are asked to determine the soil moisture class. Five soil moisture classes are considered: dry, relatively dry, moist, relatively wet, and wet. Users pick a handful of the soil sample from 5 to 10 cm depth using a spoon or scraper. Then, by referring to the visual and text guidelines shown in Table 1, they select the sample's appropriate soil moisture class.

2.2. Study area and data collection

The study area is located in the Shahriar County, Tehran province, in an area of 80 km^2 situated between E 50°58′ to 51°5′ and N 35°35′ to 35°41′. In selecting this region, factors such as soil type, land use diversity, availability of meteorological station, ease of access, and volunteers' participation were considered. The site's geomorphology is relatively uniform according to the geological maps and is classified as Young Terrace. Furthermore, the area's dominant soil texture is clay marl texture (FAO 2020) without gravel (about 80%). About 10% of it contains sandy soils with traces of volcanic rocks in the remaining. One of the selected stations is Mehrabad airport, which has a distance of 10 km from Tehran. The five-year average annual precipitation and temperature of the region are 199 mm and 16.9 °C, respectively. For developing the methodology, two sets of samplings were taken.

The first sampling was conducted on May 2, 2019. Forty-two different locations were selected.

Moreover, sand was sampled by 21 users. For selecting the sampling points, the study area was gridded into 1 km^2 cell. Then, by considering limitations such as inaccessibility of private properties, the sampling points were selected. These points were distributed among various land uses such as barren/vacant lands, farmlands, and orchards that separate EMC class as described before. Users completed the soil moisture assessment form for each sampling point. A digital image was taken from each user sample (the cell phones used had comparable digital camera specs). Besides, 42 gravimetric soil moisture measurements were collected from the same locations to determine the exact soil moisture composition. Three sampling points were marked as reference points (RF) for future studies. According to the importance of their locations and their accessibility for cross-referencing the sampling points, these points were selected if needed. RF$_1$, RF$_2$, and RF$_3$ sampling points are

Table 1 | Soil moisture classes and the provided user guide

	Classifications		
	Completely Dry	**Relatively Dry**	**Damp**
Visual Guide			
Text Guide	Soil particles do not stick together. The palm of the hand will be filled with dust as a result of kneading the soil.	Some soil particles stick together. The palm of the hand will be filled with dust as a result of kneading the soil.	Most of the soil particles stick together. Soil mass is fragile. Soil particles are visible on the hand.
	Relatively Wet	**Completely Wet**	
Visual Guide			
Text Guide	The surface of the soil is rough. Soil mass has some plasticity.	The surface of the soil is soft. Soil mass has significant plasticity.	

located next to a water withdrawal well, refinery pipeline, and gas pressure reducing station, respectively. The number of data is a limitation of this work. Further work is needed in future studies or for accurate implementation.

To analyze the collected data, the distribution of the gravimetric soil moisture data was studied by plotting the histogram and fitting a theoretical distribution function. Besides, users' performance was analyzed to evaluate the effectiveness of the developed subjective soil moisture assessment form. If the provided guidelines cannot distinguish samples with different soil moistures, the degree of mutual understanding of the users in selecting soil moisture class would be low. Therefore, users' opinion at each sampling point was compared with the data mode (as the most-selected soil moisture class) for all sample classifications at that location.

In the second sampling conducted on October 11, 2019, eight different locations were selected, and ten users carried out the data collection process. The first sampling includes that required a more thorough investigation. However, field sample collection was not needed in the second one, and only the questionnaire was used. The second dataset was reduced as this set was used for testing only and not for lab screening or model development purposes. The number of users was reduced to check whether the models could quantify the subjective data when their size is limited. Figure 2 shows the study area and the first, second, and reference sampling points.

In this study, MODIS MOD11A1 land surface temperature data was retrieved for both sampling days.

2.3. Digital image processing

In this section, digital images regarded in the data collection are processed, and the relationship between soil color and moisture is investigated. For this purpose, each soil sample's image was analyzed by the Environment of Visualizing Images (ENVI) software. The images for communicating with users and soil classifications were taken by a 12 Megapixel Canon DIGITAL IXUS 200 IS camera.

The Red-Green-Blue (RGB) model most often represents digital images. In this model, each pixel of the image contains three digital numbers (DNs). DNs are dimensionless values representing light reflectance in the red, green, and blue bands. Combinations of these DNs are used to represent various colors. The range of DNs is defined by the bit depth concept, which refers to the color information stored in an image. If an image has a higher bit depth, it could store and represent more colors. For a N-bit digital image, each DN ranges from 0 to $2^N - 1$ (for an 8-bit image, the DN could be as high as 255). As the light reflectance in 'bright' colors is more than 'dark' images, the 'bright' pixels have high DNs, while the 'dark' pixels have low DNs. In other words, '0' means no reflection, and '2^N-1' is the maximum reflection in each color band (e.g., an 8-bit image could store and illustrate 16.7 million ($2^8 \times 2^8 \times 2^8 = 2^{24}$) different colors).

In this study, the k-means clustering approach (Figure 3) was utilized to separate the soil image from its background (classify into two clusters). In the first iteration, clusters would be selected randomly. Then the cluster centers are moved to minimize the total within-cluster variance. These steps are iterated until convergence (Hastie *et al.* 2009). After clustering all data points during the first iteration, the new centroid (mean) of all data points in each cluster was calculated by assigning data points to their closest cluster centers according to the Euclidean distance function. By repeating this iterative process, the final position of clusters is determined when the same points are assigned to each cluster. Bishop (2006) showed that three more iterations are needed to converge.

2.4. Quantifying subjective data

By utilizing two ANN models in the MATLAB software, the users' subjective data were converted to quantitative soil moisture data. The first developed model (MLP) was a simple basic ANN. The second one was a radial basis function (RBF) neural network classified as a statistical ANN. For observing the application of the explanatory moisture condition data in the data-driven models, locations without vegetation such as vacant land or partially developed such as residential neighborhoods, the explanatory moisture condition were classified as EMC I. The other positions with some level of vegetation cover were categorized as EMC II. This is a different classification than the Antecedent Moisture Condition (AMC) method for estimating excess rainfall in the SCS (Soil Conservation Service) method.

The input layer variables are the users' selection of soil moisture classes, image color specifications at each sampling point, and the explanatory moisture condition of each sample. For preventing over-fitting (over-modeling) in the MPL, 250 input-target data pairs were generated and utilized in the model development based on the 42 observational data. The target data were generated randomly based on the available gravimetric soil moisture data (normal distribution). Although there
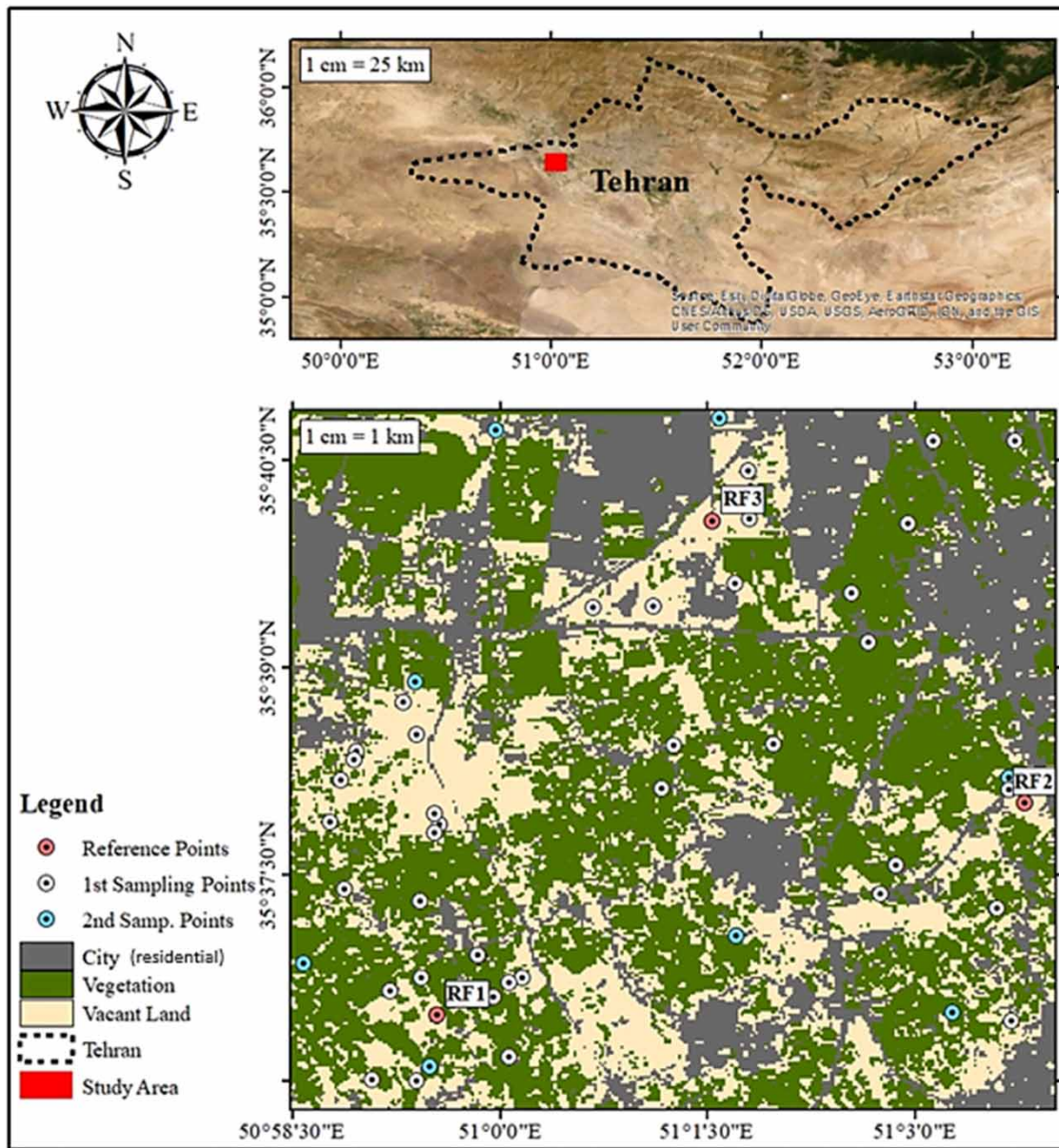
Figure 2 | Study area and sampling points located in Shahriar County, Tehran Province.

was no risk of over-fitting in the RBF, this network was developed with the same input-target datasets as the MLP so that the results would be comparable.

The architecture of the RBF network is shown in Figure 4. This network uses the Gaussian transfer function (Equation (1)) in the neurons' structure. The output of this function approaches 0–1, when the Euclidian distance between X and C_j approaches a considerable value to 0, respectively. The value of the output between those limits depends on h (known as the spread).

$$f(X, C_j) = e^{-\left[\|X-C_j\| * (0.8326)/h\right]^2} \tag{1}$$

where X is the input vector, C_j is the centroid vector of the Gaussian function of the jth neuron, and h is the spread of the Gaussian function.
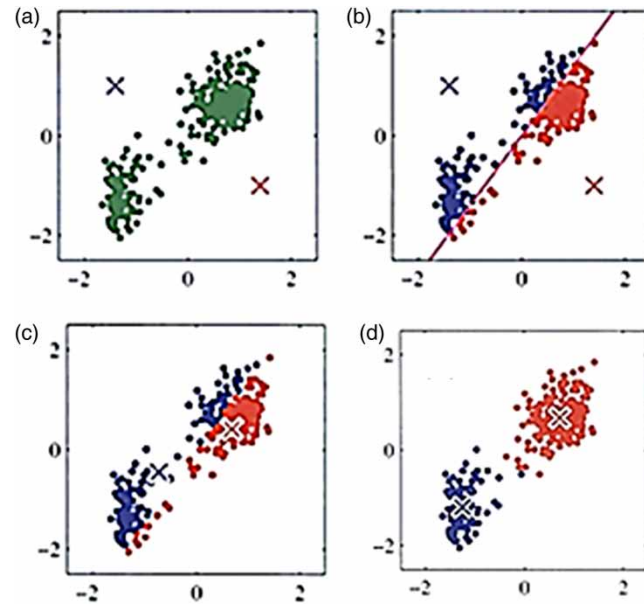
Figure 3 | An illustration of clustering of a two-component dataset into two clusters by the k-means method (Adopted from Bishop 2006).
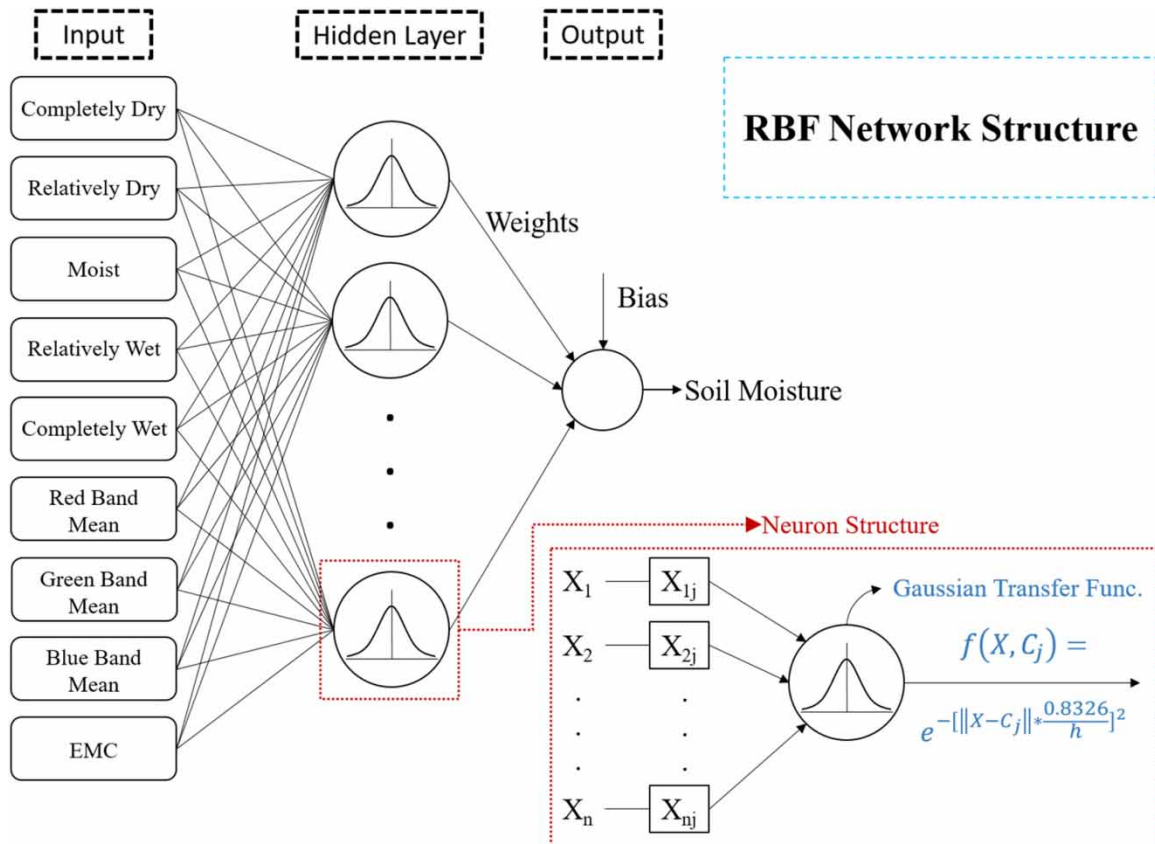

Figure 4 | The architecture of the RBF model for quantifying the soil moisture classifications.

The general form of calculating a dependent variable Y by predictor X is then:

$$Y = \sum_{j=1}^{M} W_j f(X, C_j) + b \tag{2}$$

where M is the number of neurons, W_j is the weight of connection from jth neuron to the output layer, and b bias of the output layer.

The structure of the statistical neural networks depends on the historical training dataset. For estimating the output variable for a new given input vector by these networks, the new input vector would be compared with the input vectors of the historical training dataset. As the new input vector gets closer to a particular historical input vector, the new output would get closer to that historical input's corresponding output. In RBF, classified as statistical networks, the same logic is implemented by considering the centroid (mean) of the transfer functions of the neurons as the observed historical input vectors. It contrasts with the MLPs, in which transfer functions of the neurons are log-sig and independent of the training data; the historical dataset's effect in the RBF method is directly related to the input vector. As the distance increases, the effect of that neuron on the output vector decreases and eventually approaches zero. For this purpose, 'h' is defined as the spread of the Gaussian transfer function similar to 3 standard deviations in the normal distribution. Therefore, 'h' is the influence radius for the new input vector data. Its optimal value is calculated by an iterative method. If the input vector data distance is less than 'h', the output vector is affected by that neuron.

The training procedure of the RBFs consists of two main stages. In the first stage, the hidden layers' architecture would be regulated. The connections' weights (from the hidden layer to the output layer) would be determined in the second stage. An iterative try-and-error approach was utilized to determine the number of the hidden layer's neurons and examine different spread values. Averaged simulation error for the test data was considered the indicator of the RBF network's performance.

2.5. Producing soil moisture layouts

Soil moisture layouts are crucial tools for detecting spatial variability and improving hydrological models' accuracy. Various geostatistical methods could be utilized to model the spatial variability of soil moisture. This study used the ordinary kriging (OK) technique and its variants to regionalize observed soil moisture values. The OK model uses a set of statistical tools to predict the value of a given variable (in this case, soil moisture) at locations where there is no measurement, so the predicted soil moisture at an unsampled location (point other than sampling points) is given as:

$$\text{pred}(x_0) = \sum_{i=1}^{n} w_i(x_0) \times \text{obs}(x_i) \tag{3}$$

where pred (x_0) is predicted soil moisture at the point x0, $w_i(x_0)$ is kriging weights for point x_0, obs (x_i) is observed soil moisture at the point x_i, n is the number of observation points.

OK is a development of the semivariogram, which explains how soil moisture varies with distance among the sampling locations. As shown in Equation (4), the semivariogram is half the expected squared difference between the two locations' soil moisture values. The half-squared difference between each observation pair was calculated and plotted versus their distance for developing the semivariogram. Then, an empirical semivariogram was produced by calculating the data's average in each lag interval by considering a lag distance. An automated fitting procedure was then followed for fitting the semivariogram of soil moisture using the Gaussian model.

$$\gamma(h) = \frac{1}{2} E[(z(x_i) - z(x_i + h))^2] \tag{4}$$

where $\gamma(h)$ is semivariogram value at distance h, $z(x_i)$ is the value of the variable at the point x_i, $z(x_i + h)$ is the value of the variable at a point with a distance h from point x_i. In addition to the basic version of the OK, an extension of this geostatistical method was utilized to regionalize soil moisture data. By considering a search radius, only sampling points that were located in a certain neighborhood of the estimation point were used. The search radius was selected as half of the semivariogram range, which another investigator uses. As a second extension, the influence of direction dependency was studied by utilizing

multi-directional kriging (MDK) based on Friedland *et al.* (2017). For this purpose, directional semivariograms were developed for various azimuth angles, which are the angles that a particular direction makes with the north. Then, an ellipse was sketched to determine the major and minor directional angles with the highest and the lowest spatial correlation, respectively, as shown in Figure 5. First, the range of each directional semivariogram would be plotted against its azimuth angle in a polar coordinate system (green points in Figure 5). Then, an ellipse would be fitted to these discrete points.

All data processing and analysis for OK were done by ArcGIS software, version 10.5. The soil moisture mapping accuracy was evaluated by a cross-validation scheme and calculating mean error (ME), average standard error (ASE), and RMSE indices (Yang *et al.* 2009) as follows:

$$ME = \frac{1}{n} \sum_{i=1}^{n} [\,\mathrm{pred}(x_i) - \mathrm{obs}(x_i)] \tag{5}$$

$$ASE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[\mathrm{pred}(x_i) - \left(\frac{\sum_{i=1}^{n} \mathrm{pred}(x_i)}{n} \right) \right]^2} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [\,\mathrm{pred}(x_i) - \mathrm{obs}(x_i)]^2} \tag{7}$$

where $\mathrm{obs}(x_i)$ is observed soil moisture at the point x_i, $\mathrm{pred}(x_i)$ is predicted soil moisture at the point x_i and n is the number of observation points.

2.6. Validation of citizen science with satellite data

The relationship between the soil moisture data and the land surface temperature (LST) data was investigated. LST data was retrieved and reprojected for the study area from the MODIS (Moderate Resolution Imaging Spectroradiometer) MOD11A1 product (Wan *et al.* 2015). MODIS is an imaging sensor launched into Earth orbit by NASA in 1999 onboard the Terra satellite and in 2002 onboard the Aqua satellite. MODIS is designed to provide measurements in large-scale global dynamics and collect data in 36 spectral bands ranging in wavelengths from 0.4 to 14.4 μm and varying spatial resolutions. After retrieving MODIS MOD11A1 LST data with 1 km spatial resolution (simultaneous with soil moisture data of the first sampling), each pixel was divided by the maximum LST to make it dimensionless.

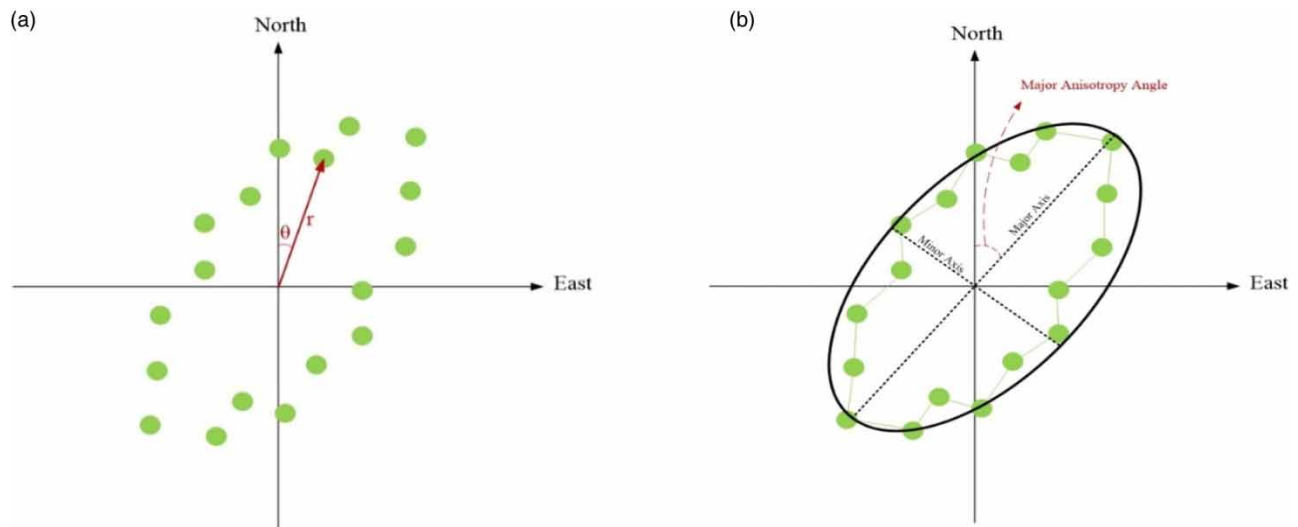$$LST_{DL(i)} = \frac{LST_{(i)}}{LST_{\max}} \tag{8}$$

Figure 5 | Elliptical representation of semivariograms in multi-directional kriging. Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/hydro.2021.029.

where $LST_{DL(i)}$ is dimensionless LST at pixel number i, $LST_{(i)}$ is LST at pixel number i and $LST_{\max}$ is the maximum observed LST. Finally, a regression model was developed to estimate soil moisture based on LST data. Linear, exponential, and power regression models were tested. The power model was selected as it yielded the best results (Equation (9)).

$$SM_{LST} = a \times LST_{DL}^{b} \tag{9}$$

where SM_{LST} is soil moisture derived from LST and a, b are regression constants.

After developing the ANN and regression models (using the first sampling data), these models were utilized to estimate soil moisture using the second sampling data in the absence of gravimetric soil moisture data. LST-based soil moisture data were used to validate the data collected by the users in real-time. After comparing these datasets, the difference between them was quantified to depict the error of citizen science-based data.

2.7. Data estimation platform – error analysis

After developing the ANN-based and LST-based soil moisture models, they should be implemented for any new real-time sampling to cross-validate the results. The cross-validation criterion for the validity of the results from real-time sampling was the absolute error between ANN-based and LST-based soil moisture at each point of the real-time sampling in the range of mean plus/minus the standard deviation of the modeling stage in its EMC group. The results from ANN-based and LST-based models for real-time sampling were modified and combined in the next step. For this purpose, first, due to the importance of the location of points, the study area was divided into polygons using the Thiessen method for the real-time sampling points. Then two correction factors were calculated based on errors of the results. First, the difference between the field data and the ANN-based soil moisture in each first sampling point in the polygon was estimated to calculate the error, reporting their mean as error$_1$. Next, error$_2$ was calculated as the difference between the field data and LST-based soil moisture in each point of first sampling in each polygon. Then, a correction factor was calculated based on the inverse ratio of errors in each polygon. Then the correction factor 1 ($error_2/error_1 + error_2$) related to the estimation from ANN-based model was multiplied by the ANN-based soil moisture and also the correction factor 2 ($error_1/error_1 + error_2$), related to the estimation from the LST-based model was multiplied by the LST-based soil moisture. Next, the modified values are added together. The estimated real-time data was obtained for each point of real-time sampling. Finally, for regionalizing the soil moisture data, the average grid value with min and max cell/grid values of soil moisture could be calculated for the study area. The average value's representative location is a virtual station at the centroid of the study area's pervious part (not counting roads, buildings, and other impervious sections). This location is crucial if we want to repeat this process to develop soil moisture time series.

3. RESULTS AND DISCUSSION

Based on the proposed methodology, considering the data analysis as the first step is followed by the image processing outcomes. The data-driven models and regionalizing soil moisture are discussed followed by cross-validation of citizen science with satellite data. Finally, the platform for soil moisture data estimation is investigated and discussed.

3.1. Data analysis

The collected data was analyzed to study the distribution of the gravimetric soil moisture data and the users' performance. The data distribution was studied to determine the best theoretical distribution function that fits the data to generate extra data pairs for developing the data-driven models. Although the normality of the data is not a prerequisite for utilizing the kriging method, it could increase the accuracy of the ordinary kriging technique (Bagheri Bodaghabadi 2018). Results show a close match between the observed data and the fitted normal distribution. The Kolmogorov-Smirnov test was conducted, and the normality of the data was accepted at a 5% significance level.

The users' performance analysis indicated that about 65% of the users' soil classification choices were compatible with the mode at each sampling point. In this study, the number of choices agreed or were off among subscribers with only one wetness classification (very dry, dry, etc.) was about 95%. These results are similar to Rinderer et al. (2015). In that paper, the users' agreement in soil moisture subjective classifications was 91 and 93% in two different sampling groups.

3.2. Image processing

For investigating the correlation between soil color and moisture, digital image processing was conducted. In the first step, each sampling point's image was classified into two classes (soil image and its background) utilizing a k-means unsupervised classification scheme through an iterative approach. Results of implementing the k-means clustering for one of the samples which were collected during the first sampling (selected randomly) are presented in Figure 6. Each cluster's pixels were analyzed. The frequency of the pixels with different digital numbers (DNs) was plotted in each color band. Background pixels were removed to proper soil moisture estimation (Figure 6(c) and 6(d)). Then the average DN of the red, green, and blue bands was calculated for each image's soil cluster pixels, and the correlation of the average DNs with the soil moisture was investigated.

As shown in Figure 7, the average DN declines in all three bands by increasing soil moisture. This decrease is due to the fact that water reflectance is smaller than soil reflectance in visible spectra. By increasing the water content, the reflectance of the sample decreases as well as the average DN. By detecting these alterations via utilizing proposed image processing
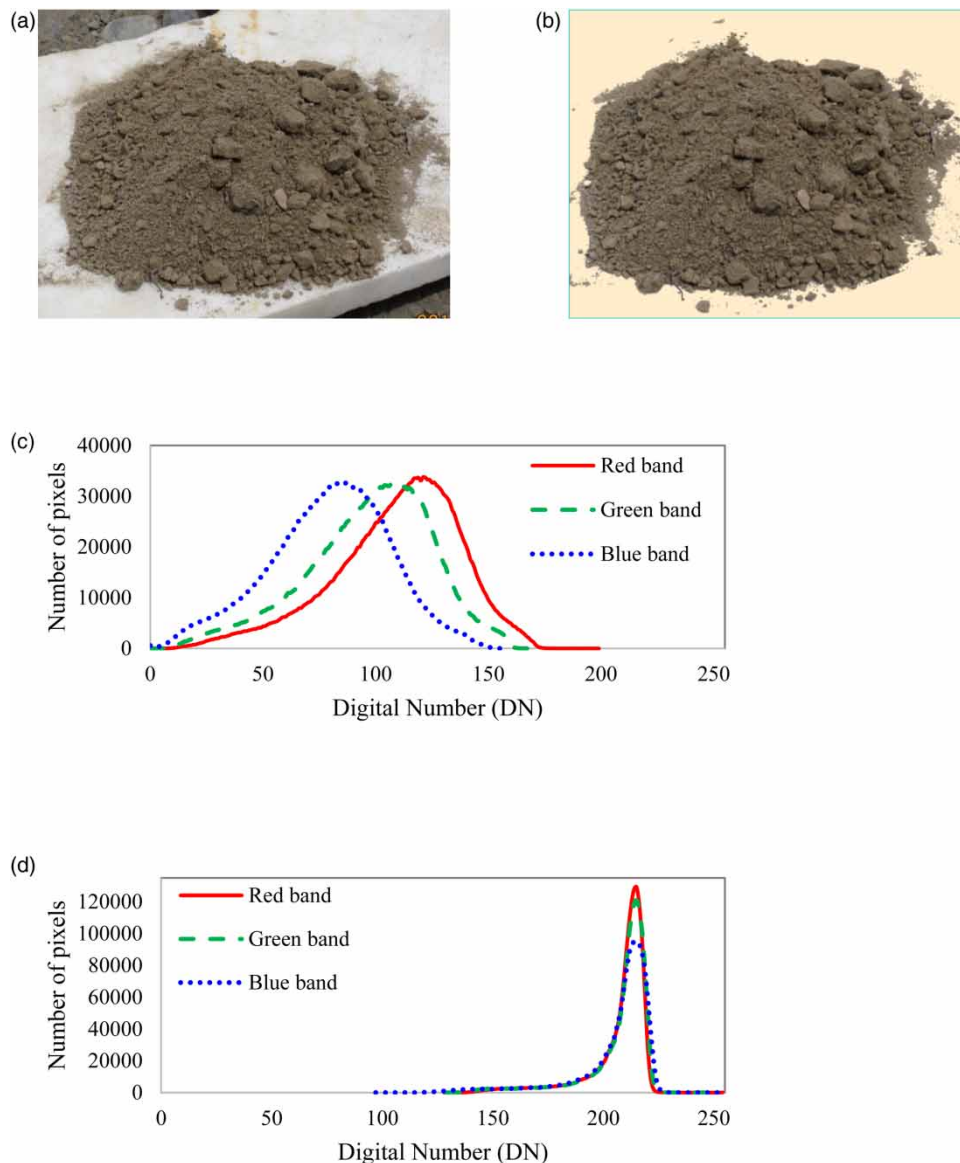

Figure 6 | Results of k-means classification for a random sample: (a) unclassified image (b) classified image; No. of pixels in each color band for (c) soil cluster (d) background cluster.
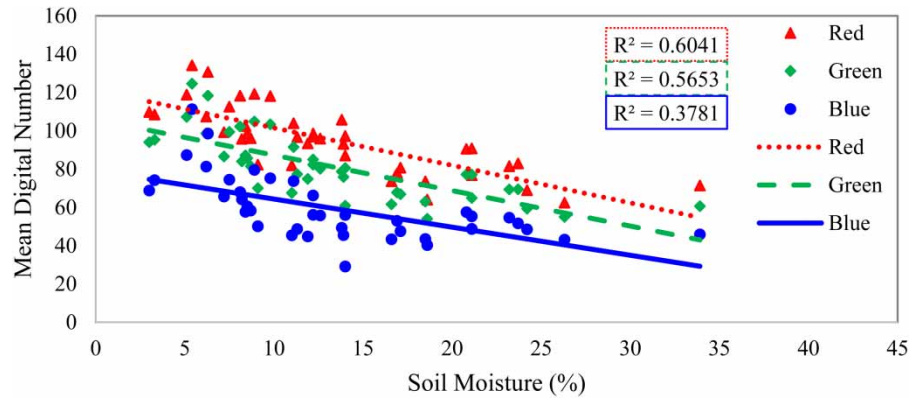
Figure 7 | Average Digital Number (DN) of each color band for soil moisture of the first sampling.

techniques, soil moisture variations could be captured. R^2 between the average DN of the red, green, and blue band and soil moisture was calculated to be 0.60, 0.57, and 0.38, respectively. These values show the high potential of soil color in estimating soil moisture. As a result, they were used in the structure of the ANN model as predictors. The average DN of red, green, and blue color bands for the image of each soil sample were utilized as a part of the ANN input vector to estimate soil moisture for that sample.

3.3. Data-driven modeling

Both MLP and RBF models had nine inputs and one output. The first five inputs consisted of the proportion of the users' selection of the soil moisture class at each sampling point. The sixth to eighth inputs were the red, green, and blue bands average DN. The last input was dedicated to the EMC classification of the sample. The target values of the models were the gravimetric soil moisture data. One of the Inputs-target pairs is presented in Table 2. At this point, 19, 57, and 24% of the users have chosen completely dry (CD), relatively dry (RD), and damp (DD) classes, respectively. None of the users have chosen relatively wet (RW) and completely wet (CW) classes at this sampling point. The sixth to eighth inputs are 109.72, 94.01, and 68.77, respectively, and the EMC class is labeled as EMC I. In the last column of the table, the target value is demonstrated.

For developing the MLP model, 85 and 15% of the data were used for training and testing, respectively. The model was built and tested for ten randomly selected different training-testing datasets. Several network architectures were designed, and their performance was evaluated by statistical indicators such as R^2 and RMSE (Table 3). The best results were obtained from an MLP network with six neurons in the hidden layer as the differences were minimal. Therefore, this architecture was chosen as the best MLP model. The optimum MLP model was quite efficient in predicting soil moisture, as its R^2_{Overal} (R^2_{Train}, R^2_{Test}) was 0.95 (0.96, 0.93), and its $RMSE_{Overal}$ ($RMSE_{Train}$, $RMSE_{Test}$) was 1.33% (1.28%, 1.48%).

For developing the RBF network, various data ratios (from 5% to 95% with 5% increments) were selected to train the model, and the optimum ratio was determined. Different spread values, from 0.1 to 5.0 with 0.1 incremental steps, were examined for each ratio. The optimum spread is the spread in which the minimum error occurs (Figure 8). By increasing the percentage of the data used in the training stage, the mean absolute error (MAE) decreased. The minimum MAE, which was 2.7%, occurred when 85% of the data was used in the training stage, and the corresponding optimum spread was 2.9% (Figure 8(a) and 8(b)).

For comparing the MLP and RBF networks' performance, although the MLP model yielded lower errors, the benefits of utilizing the RBF network are undeniable. Contrary to MLPs, RBF networks could be utilized for data-driven modeling
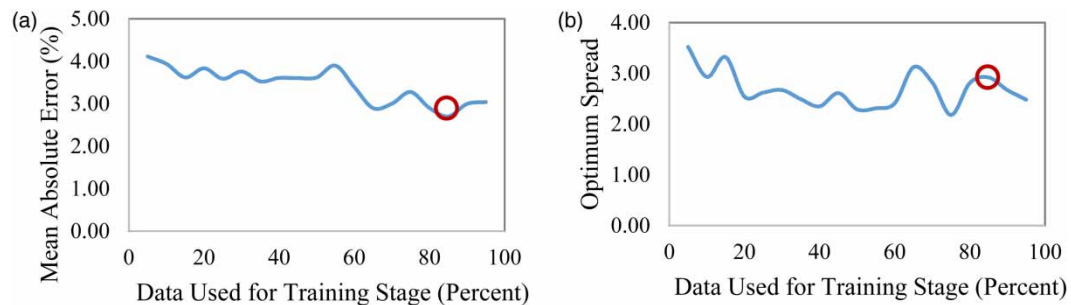
Table 2 | Example of inputs-target of the ANN models for a sampling point

CD	RD	DD	RW	CW	Red	Green	Blue	EMC	Gravimetric Soil Moisture (%)
0.19	0.57	0.24	0.00	0.00	109.72	94.01	68.77	I	3.0

CD, completely dry; RD, relatively dry; D, damp; RW, relatively wet; CW, completely wet.

Table 3 | Statistics of different ANN neurons structures

No. of Neurons	R^2_{Train}-R^2_{Test}[a]	$RMSE_{Train}$- $RMSE_{Test}$ (%)[a]
6	0.02	0.20
8	0.04	0.49
10	0.02	0.37
12	0.03	0.41
14	0.05	0.63

[a]Absolute values.

Figure 8 | RBF network testing results of the data used for training stage: (a) MSE (b) optimum spread, 85% marked.

when the number of observations (size of the training dataset) is limited. Besides, the error of the RBFs would approach zero by increasing the number of observational data. These neural networks could estimate soil moisture in other sampling points with the same soil texture. However, the model should be recalibrated to use the proposed method in a region with different geologic structures.

3.4. Regionalizing soil moisture

Semivariograms were presented in Figure 9(a) and 9(b) to regionalize gravimetrically and quantified soil moisture data by the OK method. The Gaussian model was selected to model the spatial variability of soil moisture in both datasets. Depending on the criteria suggested by Cambardella *et al.* (1994), the spatial autocorrelation is considered as high if the nugget to sill ratio is less than 0.25, to be moderate if between 0.25 and 0.75, and to be low if higher than 0.75 (Dall'Agnol *et al.* 2020). In this study, the fitted semivariograms have nugget to sill ratios of 0.32 and 0.44 for gravimetric and quantified soil moisture data, respectively, demonstrating a moderate (close to high) spatial dependence in both cases. Compared to similar studies, Huaxing *et al.* (2009) results are consistent with this study. As far as spatial dependency is concerned, this paper has used moderate spatial autocorrelation, and Huaxing *et al.* (2009) used a moderate spatial dependency index.

For producing soil moisture layouts by the MDK, directional semivariograms were plotted for 0° to 360° azimuth angles with 10° incremental steps. The azimuth angle is the angle that a particular direction makes with the north direction. A directional ellipse was sketched by comparing the plotted semivariograms, and major and minor directional angles were determined. These angles show the directions in which data have the highest and lowest spatial correlation, respectively. The fitted expression is as follows in which x and y are Cartesian coordinates:

$$1.25x^2 - 1.46xy + 1.41y^2 = 0.00123 \tag{10}$$

The semivariograms of the major directional angle have the most extensive range compared to the other directions' semivariograms (Figure 9(c) and 9(d)).

After developing the semivariograms, soil moisture layouts were derived from the gravimetric and quantified soil moisture data. For evaluating the accuracy of the utilized geostatistical methods, statistical indices such as ASE, RMSE, and ME were calculated (Table 4).
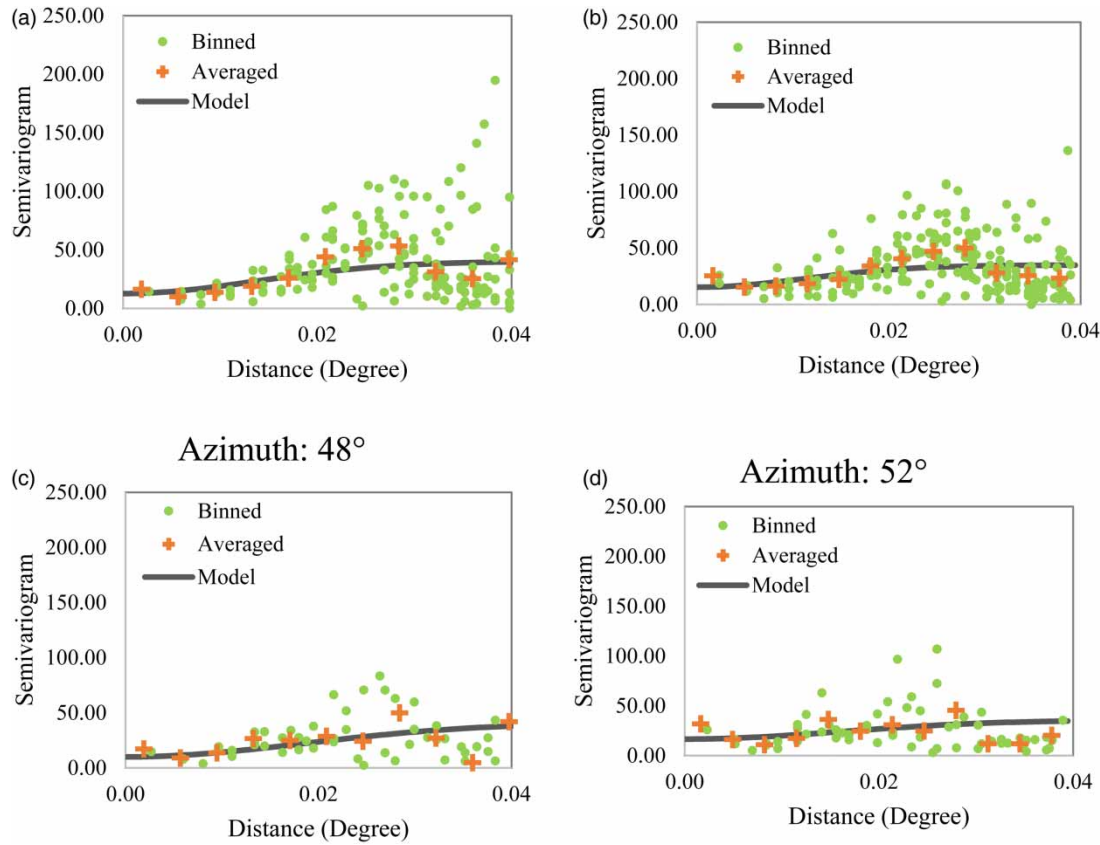
Figure 9 | Fitted Gaussian semivariograms for gravimetric and quantified soil moisture data: OK (a, b) and MDK (c, d) methods.

Table 4 | Statistical error indices of kriging techniques

Data type	Method	ME (%)	ASE (%)	RMSE (%)
Gravimetric	OK	−0.097	4.48	5.05
	MDK	−0.003	4.23	4.80
Quantified	OK	−0.064	4.83	5.23
	MDK	−0.022	4.89	4.94

According to Table 4, the soil moisture layouts' precision has increased by utilizing MDK instead of the basic OK. Methods OK, and MDK yielded MEs of −0.097%, −0.018%, and −0.003% for gravimetric data and −0.064%, −0.047%, and −0.022% for quantified data respectively. These values revealed the advantage of the MDK with the smallest ME and the reduced RMSE (40.8% for the gravimetric data). So, the MDK was chosen as a better method for producing soil moisture layouts. Figure 10(a) and 10(b) represent soil moisture layouts derived from the gravimetric and quantified soil moisture data by the MDK method. Besides, the first sampling points are presented in these layouts.

As shown in Figure 10, there is a southwestern-northeastern direction in the soil moisture layouts for both gravimetric and quantified data. Besides, according to the results, soil moisture and land use patterns are compatible. The soil moisture value is much more in the vegetated lands and less in barren/vacant lands. Layouts provided by the kriging technique could be utilized to study the spatial variations of soil moisture. This technique is further utilized in the 'Soil Moisture Data Estimation Platform' section to provide a real-time soil moisture layout and calculate the study area's average soil moisture.

3.5. Cross-validation of citizen science with satellite data

In order to investigate the relationship between the soil moisture and land surface temperature layouts, LST data was retrieved from the MODIS MOD11A1 product. MODIS sensor collects data in 36 spectral bands and various spatial
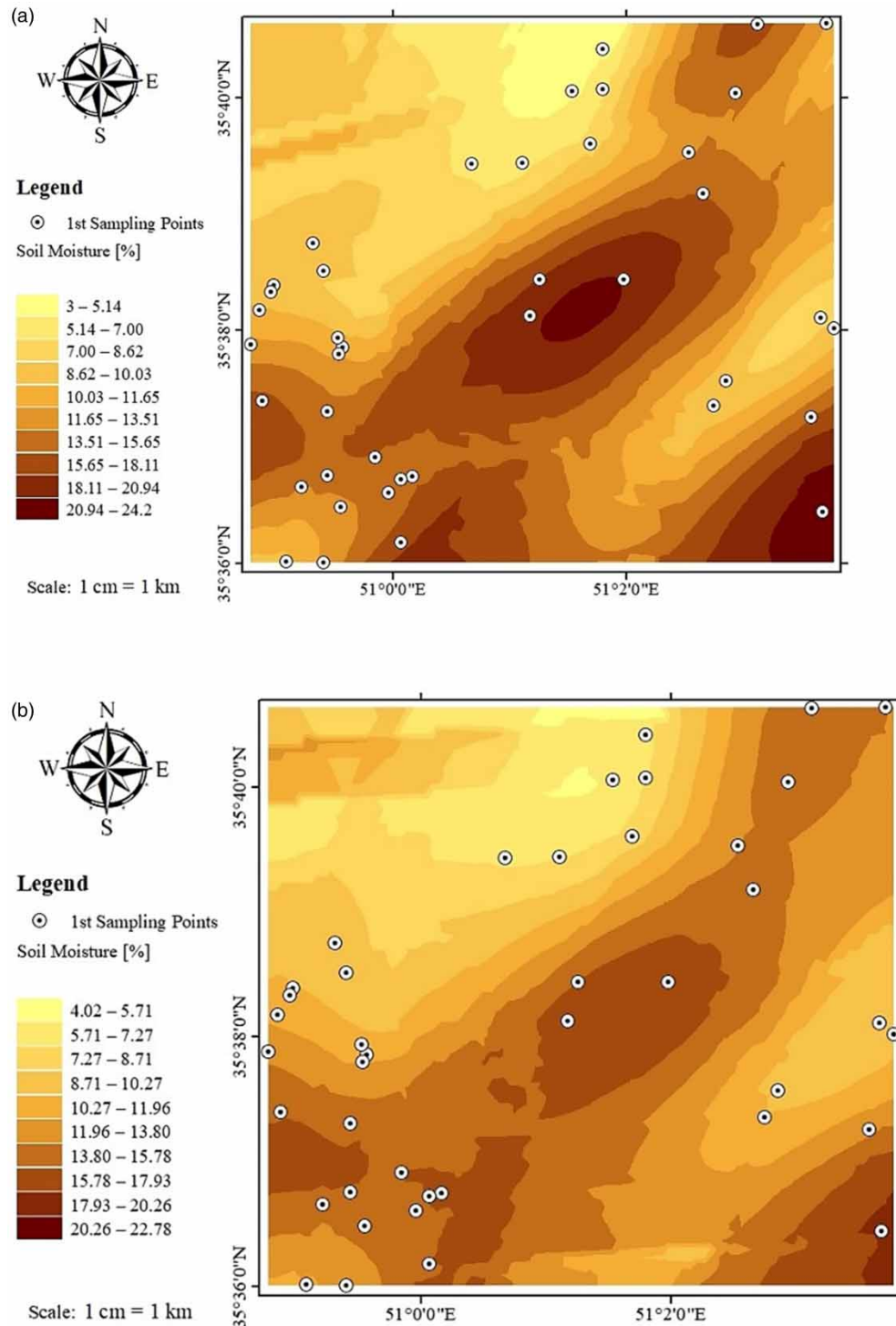
Figure 10 | Soil moisture layout generated from (a) gravimetric data and (b) quantified data by multi-directional kriging (MDK) technique.

resolutions. MOD11A1 product developed based on the Terra MODIS data provides 1 km^2 land surface daily temperature data. Then, the spatial resolution of LST data was increased to 0.04 km^2 by interpolating the original 1 km^2 data over a 200 m $\times$ 200 m grid by utilizing the MDK. Figure 11 shows the high-resolution LST layout for the first sampling date (May 2, 2019).

About 80% of the data were used to develop a power function correlation between LST and soil moisture data. The rest were used to test the relationship. Ten different building-testing datasets were selected to develop and test the relationship
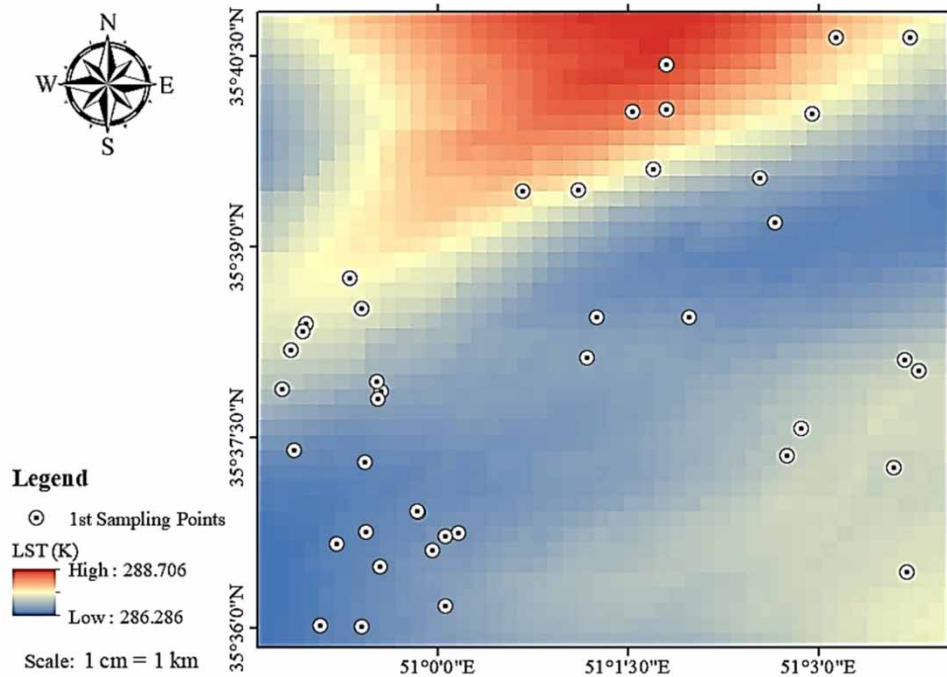
Figure 11 | High-resolution land surface temperature (LST) layout for the first sampling date.

to prevent over-fitting. For this purpose, ten random datasets with 80% for the building stage and 20% for the testing stage were selected, and the power relationship was developed/tested. Finally, the average R^2 and RMSE were calculated based on all datasets.

Several linear ($R^2 = 0.65$), exponential ($R^2 = 0.63$), and power relationships were tested to investigate the correlation between soil moisture and dimensionless land surface temperature. Finally, the power model was selected as the best model as:

$$SM_{LST} = 4.296 \times LST_{DL}^{-178.67} \tag{11}$$

The power relationship shown in Figure 12 yielded an average R^2_{Train} (R^2_{Test}) of 0.69 (0.68) over ten building-testing datasets, and its average $RMSE_{Train}$ ($RMSE_{Test}$) was 2.31% (2.30%).
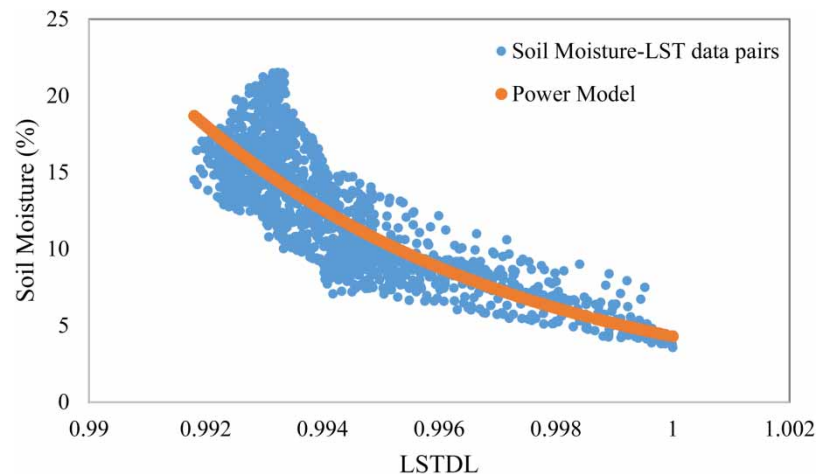

Figure 12 | Fitted power model and the soil moisture-LST data pairs.

3.6. A platform for soil moisture data estimation

The proposed method for eight points has been implemented as a pilot. First, LST_{DL} data was retrieved for the second sampling date (October 11, 2019) and converted to SM_{LST} by Equation (10). Then, the second citizen science-based dataset was quantified by the ANN model. The ANN model's error was calculated by comparing the results with SM_{LST} obtained from the regression model. Results show that quantified soil moisture data had a high correlation with SM_{LST} ($R^2 = 0.90$) and a low RMSE of 3.09%.

The absolute error between ANN-based and LST-based soil moisture data for 42 sampling points in the first sampling was calculated. Figure 13 demonstrates the frequency of the absolute error between ANN-based and LST-based soil moisture for all EMC I and EMC II points. More than 40% of all data had less than two percent error. Out of 40%, 25% belonged to the EMC I class and 15% to the EMC II class. Overall, samples taken from EMC I location displayed less error.

The mean and standard deviation of the absolute error between ANN-based and LST-based soil moisture for the first sampling and real-time sampling are shown in Tables 5 and 6. The results of the two models are more different in the EMC II class, so the mean of EMC II points is more than EMC I points. The EMC II points' standard deviation is more than EMC I point, which shows more uncertainty in the EMC II points. This indicates that when the area is in wet condition, the satellite data cannot accurately capture the region's environmental condition. Thus, the results from the citizen science-based framework are more reliable than the satellite-based model. As a result, EMC provides valuable information that is not available in satellite data.
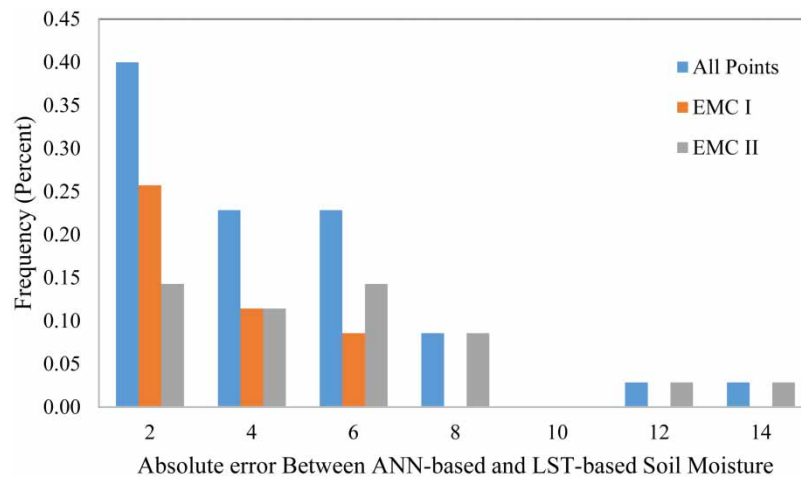

Figure 13 | Frequency of absolute error between ANN-based and LST-based soil moisture.

Table 5 | The mean and standard deviation of absolute error for the first sampling

First sampling	All points	EMC I point	EMC II points
Mean	3.46	2.33	4.41
Standard Deviation	2.96	1.64	3.45

Table 6 | The mean and standard deviation of absolute error for real-time sampling

Real-time sampling	All points	EMC I points	EMC II points
Mean	2.42	1.28	3.56
Standard Deviation	1.91	0.51	2.11

The EMC group should first be determined for each new sampling point. Then soil moisture should be estimated using the ANN and regression models, supposing the absolute error between ANN and LST-based models, in the range of mean plus/minus the standard deviation of the first sampling points' errors. For instance, for an EMC I group point in the second sampling (point C), estimation of soil moisture is validated, as the absolute error of ANN-based and LST-based soil moistures is 0.8% which is between 2.33 ± 1.64% (mean ± one std. dev.). Table 7 demonstrates the results of cross-validation for all eight points.

For modifying and combining the results from the ANN-based and LST-based models for real-time sampling, the area was divided into eight sub-regions, with the help of the second sampling points and the Thiessen polygon method (Figure 14). Part of 42 points was placed in each of the eight sub-regions, and the errors were analyzed.

Table 7 | Real-time sampling data points information

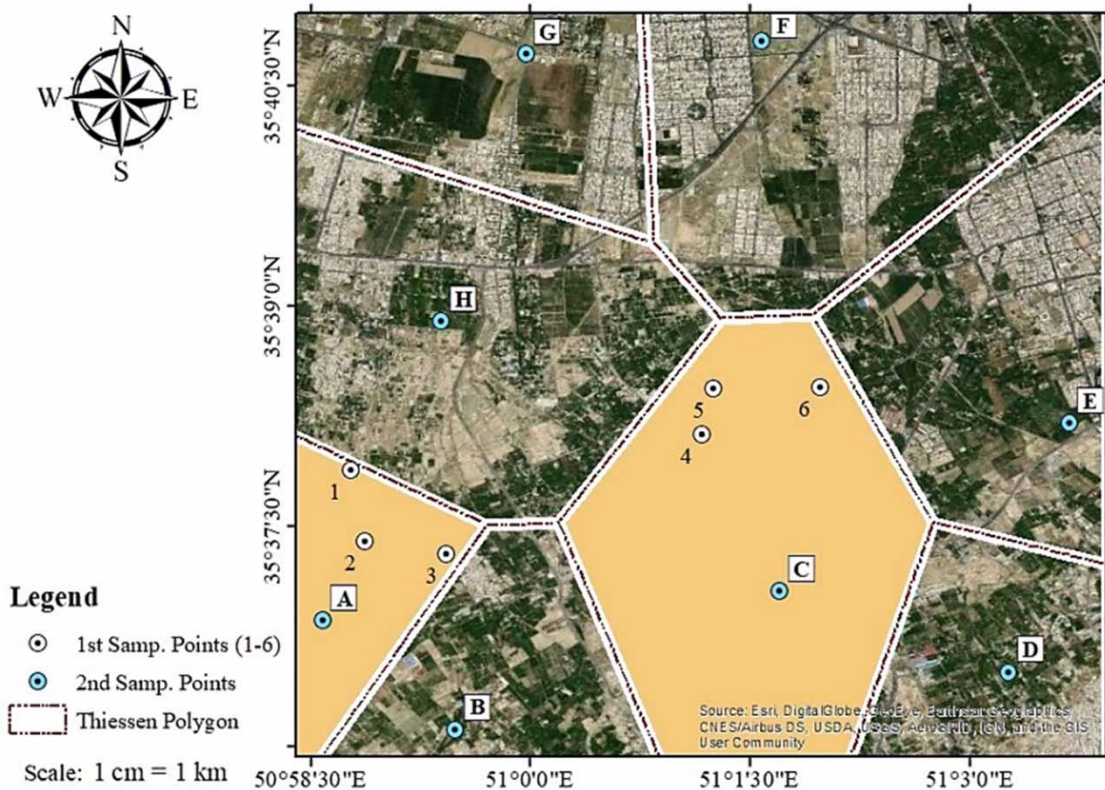| No.	EMC	Soil moisture (%) ANN-based	LST-based	Absolute error	[Mean+ Std.Dev, Mean − Std.Dev]	Cross-validation criterion
A	II	22.86	18.71	4.15	[0.96, 7.86]	Pass
B	II	22.24	15.79	6.45	[0.96, 7.86]	Pass
C	I	10.32	11.13	0.80	[0.7, 3.97]	Pass
D	II	10.82	10.25	0.57	[0.96, 7.86]	Pass
E	II	12.30	9.23	4.15	[0.96, 7.86]	Pass
F	I	6.36	4.68	1.69	[0.7, 3.97]	Pass
G	I	5.06	6.94	1.88	[0.7, 3.97]	Pass
H	I	10.82	10.08	0.75	[0.7, 3.97]	Pass

Figure 14 | Selection of eight sub-regions with Thiessen polygons and representative.

For example, in point A, errors of points 1, 2, and 3 were calculated and averaged for both ANN-based and LST-based data for this subregion. Then the correction factors for ANN-based and LST-based data were calculated based on the inverse error ratio. The mean of $error_1$ and $error_2$ for point A was 1.64% and 2.89%. The calculated errors were then converted to the correction factors ranging from 0.17 to 0.91 and regarded in soil moisture estimation. The errors and correction factors for all eight points and the final adjusted soil moisture are shown in Table 8.

Correction factor1 (which is related to the estimation from ANN-based model) and Correction factor2 (which is related to the estimation from the LST-based model) range in 0.6–0.83, and 0.17–0.4. The points C and D represent the maximum error values in both ANN and LST-based models, this may be a result of the points' properties.

Then, the kriging technique was applied to the eight modified points to produce a real-time soil moisture layout with 200 m spatial resolution (Figure 15). Since eight points were not enough to use MDK, ordinary kriging (OK) was utilized to map soil

Table 8 | Modified soil moisture of real-time sampling points

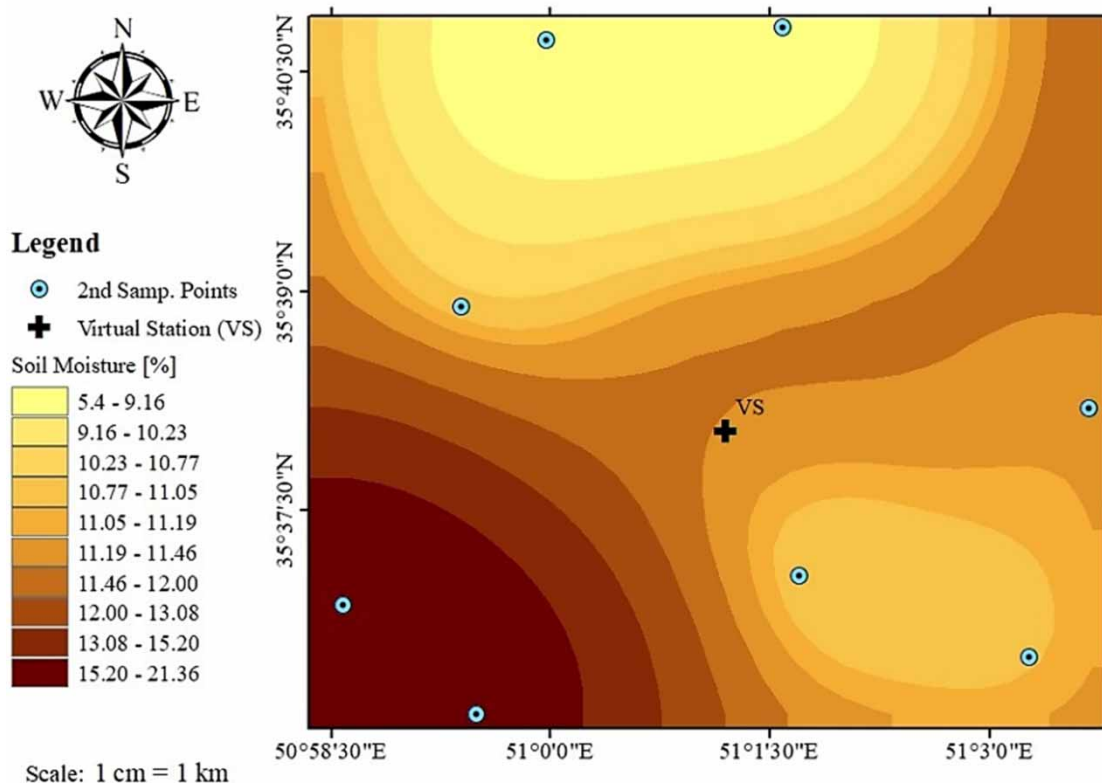| No.	ANN-based SM (%)	E_1	Correction factor$_1$	LST-based SM (%)	E_2	Correction factor$_2$	Adjusted SM (%)
A	22.86	1.64	0.64	18.71	2.89	0.36	21.36
B	22.24	1.51	0.69	15.79	3.28	0.31	20.21
C	10.32	2.56	0.71	11.13	6.36	0.29	10.55
D	10.82	2.01	0.83	10.25	9.75	0.17	10.73
E	12.30	0.86	0.60	9.23	1.31	0.40	11.09
F	6.36	0.91	0.79	4.68	3.51	0.21	6.02
G	5.06	0.50	0.82	6.94	2.22	0.18	5.40
H	10.82	0.76	0.69	10.08	1.65	0.31	10.59

Figure 15 | An example of spatial real-time soil moisture sampling for the virtual station.

moisture variations. Finally, average, minimum, and maximum cell values of soil moisture were calculated as 11.59%, 5.40%, and 21.36%, respectively.

The centroid of the study area (marked VS in Figure 15) was calculated by excluding the impermeable regions. Therefore, this point with 51° 1′ 11.64″N and 35° 38′ 2.58″E coordinates is considered a virtual station for the study area that soil moisture time series could be collected following the outlined procedure. By utilizing the proposed method, the amount of soil moisture and its range of variations at any time can be obtained for the study area to overcome the data scarcity of this vital information. As a final note related to the results, we should also reiterate that the results from the citizen science-based framework with EMC explanatory data seems to be more reliable than the satellite-based model. Therefore, EMC along with other attributes of the study provides valuable information that is not available in satellite data.

4. CONCLUSION

Soil moisture is one of the most critical hydrological variables that connect many hydrologic cycle attributes. Besides, as an indicator of climate change's long-term effects, soil moisture variation is the center of much attention. Despite the great importance of soil moisture data, it is not widely available because its sampling is relatively expensive and does not have a custodian in many developing regions. As an effort to overcome soil moisture data scarcity, this research has proposed a new framework and a platform to demonstrate citizen science's potential and capacity to collect subjective soil moisture data and land surface-based satellite data to estimate real-time soil moisture data. A pilot was selected in Shahriar County, Tehran Province. First, a user interface and a sampling process were designed to collect and analyze the subjective and gravimetric soil moisture data. The first sampling process results revealed an acceptable performance of subjective data with a high agreement among users in selecting soil moisture subjective classes. The ANN-based models were used for quantifying subjective data. For regionalizing soil moisture, layouts were produced by utilizing kriging techniques. In order to better demonstrate the contribution, scientific values, and applicability of the paper, the following concluding remarks are made. It should be reiterated that perhaps this is the first study that has explored the application of Citizen Science (social media) in estimating soil moisture for regions that have scarce soil moisture data. The application of multi-directional kriging (MDK) along with the introduction of explanatory soil moisture condition, and a proposed algorithm were utilized in this paper for collecting subjective soil moisture data at selected ungagged locations. MODIS MOD11A1 dataset was used, and a framework for error analysis was developed. AA platform was also designed to combine real-time social media (citizen science) data and satellite LST based on eight virtual substations. It provided cell values of soil moisture over 200-meter resolution kriging maps for the study area. In this way, a range of soil moisture values can be estimated on a timely (daily, etc.) basis to overcome the data scarcity of this vital information. The proposed methodology and platform can be utilized in other developing regions.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

AppEAR: citizen science in aquatic environments [WWW Document]. Available from: https://app-ear.com.ar/

Bagheri Bodaghabadi, M. 2018 Is it necessarily a normally distributed data for kriging? A case study: soil salinity map of Ghahab area, central Iran. *Desert* **23**, 284–293.

Bishop, C. M. 2006 *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F. & Konopka, A. E. 1994 Field-scale variability of soil properties in central Iowa soils. *Soil Science Society of America Journal* **58**, 1501–1511.

Crowdwater [WWW Document], Available from: https://crowdwater.ch/en/welcome-to-crowdwater/

Dall'Agnol, R. W., Michelon, G. K., Bazzi, C. L., Magalhães, P. S. G., de Souza, E. G., Betzek, N. M. & Sobjak, R. 2020 Web applications for spatial analyses and thematic map generation. *Computers and Electronics in Agriculture* **172**, 105374.

Dobriyal, P., Qureshi, A., Badola, R. & Hussain, S. A. 2012 A review of the methods available for estimating soil moisture and its implications for water resource management. *Journal of Hydrology* **458–459**, 110–117. https://doi.org/10.1016/j.jhydrol.2012.06.021

Dripps, W. R. & Bradbury, K. R. 2007 A simple daily soil–water balance model for estimating the spatial and temporal distribution of groundwater recharge in temperate humid areas. *Hydrogeology Journal* **15** (3), 433–444.

Entekhabi, D., Njoku, N., O'Neill, P., Spencer, M., Jackson, T., Entin, J., Im, E. & Kellogg, K. 2010 The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE* **98**, 704–716.

FAO 2020 Available from: www.fao.org

Floodcrowd [WWW Document], Available from: http://www.citizensciencecenter.com/assessing-flood-risks-crowdsourcing-floodcrowd/

Friedland, C. J., Joyner, T. A., Massarra, C., Rohli, R. V., Treviño, A. M., Ghosh, S., Huyck, C. & Weatherhead, M. 2017 Isotropic and anisotropic kriging approaches for interpolating surface-level wind speeds across large, geographically diverse regions. *Geomatics, Natural Hazards And Risk* **8**, 207–224. https://doi.org/10.1080/19475705.2016.1185749

García, G. A., Venturini, V., Brogioni, M., Walker, E. & Rodríguez, L. 2019 Soil moisture estimation over flat lands in the Argentinian Pampas region using Sentinel-1A data and non-parametric methods. *International Journal of Remote Sensing. Taylor & Francis.* https://doi.org/10.1080/01431161.2018.1552813

Han, P., Dong, D., Zhao, X., Jiao, L. & Lang, Y. 2016 A smartphone-based soil color sensor: for soil type classification. *Computers and Electronics in Agriculture* **123**, 232–241. https://doi.org/10.1016/j.compag.2016.02.024

Hastie, T., Tibshirani, R. & Friedman, J. 2009 The elements of statistical learning the elements of statistical learningData mining, inference, and prediction, second edition. *Springer Series in Statistics.* https://doi.org/10.1007/978-0-387-84858-7

Heil, J., Häring, V., Marschner, B. & Stumpe, B. 2019 Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: a case study with West African soils. *Geoderma* **337**, 11–21.

Huaxing, B., Xiaoyin, L., Xin, L., Mengxia, G. U. O. & Jun, L. I. 2009 A case study of spatial heterogeneity of soil moisture in the Loess Plateau, western China: a geostatistical approach. *International Journal of Sediment Research* **24** (1), 63–73.

Karamouz, M. & Fereshtehpour, M. 2019 Modeling DEM errors in coastal flood inundation and damages: a spatial nonstationary approach. *Water Resources Research* **55** (8), 6606–6624.

Karamouz, M., Nazif, S. & Falahi, M. 2012 *Hydrology and Hydroclimatology. Principles and Applications.* CRC Press, California, USA, 1466512202.

Karamouz, M., Ghomlaghi, A., Alipour, R. S., Nazari, M. & Fereshtehpour, M. 2019 Emerging and innovative technologies. IN: World Environmental and Water Resources Congress 2019. American Society of Civil Engineers, Reston, VA, pp. 85–95. https://doi.org/10.1061/9780784482322.009

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J. & Berger, M. 2001 Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing* **39**, 1729–1735.

Khadim, F. K., Dokou, Z., Lazin, R., Moges, S., Bagtzoglou, A. C. & Anagnostou, E. 2020 Groundwater modeling in data scarce aquifers: the case of Gilgel-Abay, Upper Blue Nile, Ethiopia. *Journal of Hydrology* **590**, 125214.

Kim, H. & Choi, M. 2015 Impact of soil moisture on dust outbreaks in East Asia: using satellite and assimilation data. *Geophysical Research Letters* **42**, 2789–2796. https://doi.org/10.1002/2015GL063325

MEA (Millennium Ecosystem Assessment) 2005 *Ecosystems and Human WellBeing: Current State and Trends: Findings of the Condition and Trends Working Group.* Island Press, Washington, DC.

Njue, N., Kroese, J. S., Gräf, J., Jacobs, S. R., Weeser, B., Breuer, L. & Rufino, M. C. 2019 Citizen science in hydrological monitoring and ecosystem services management: state of the art and future prospects. *Science of the Total Environment* **693**, 133531.

Porsline: online survey software and tools [WWW Document], Available from: https://porsline.ir/

Rinderer, M., Kollegger, A., Fischer, B. M. C., Stähli, M. & Seibert, J. 2012 Sensing with boots and trousers – qualitative field observations of shallow soil moisture patterns. *Hydrological Processes* **26**, 4112–4120. https://doi.org/10.1002/hyp.9531

Rinderer, M., Komakech, H. C., Müller, D., Wiesenberg, G. L. B. & Seibert, J. 2015 Qualitative soil moisture assessment in semi-arid Africa – the role of experience and training on inter-rater reliability. *Hydrology and Earth System Sciences* **19**, 3505–3516. https://doi.org/10.5194/hess-19-3505-2015

Rossiter, D. G., Liu, J., Carlisle, S. & Zhu, A. X. 2015 Can citizen science assist digital soil mapping? *Geoderma* **259–260**, 71–80. https://doi.org/10.1016/j.geoderma.2015.05.006

Sandholt, I., Rasmussen, K. & Andersen, J. 2002 A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment* **79**, 213–224. [CrossRef].

Sulieman, M. M. & Algarni, A. M. 2019 Soil organic carbon mapping and prediction based on depth intervals using kriging technique: a case of study in alluvial soil from Sudan. *Eurasian Journal of Soil Science.* https://doi.org/10.18393/ejss.492466

Sun, L., Sun, R., Li, X., Liang, S. & Zhang, R. 2012 Monitoring surface soil moisture status based on remotely sensed surface temperature and vegetation index information. *Agricultural and Forest Meteorology* **166**, 175–187.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M. & Traver, I. N. 2012 GMES Sentinel-1 mission. *Remote Sensing of Environment* **120**, 9–24.

Wan, Z., Hook, S. & Hulley, G. 2015 MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1 km SIN Grid V006. NASA EOSDIS L. Process. DAAC. https://doi.org/10.5067/MODIS/MOD11A1.006

Wang, L. & Qu, J. J. 2009 Satellite remote sensing applications for surface soil moisture monitoring: a review. *Frontiers of Earth Science in China* **3**, 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wilson, D. J., Western, A. W., Grayson, R. B., Berg, A. A., Lear, M. S., Rodell, M., Famiglietti, J. S., Woods, R. A. & McMahon, T. A. 2003 Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand. *Journal of Hydrology* **276**, 254–274. https://doi.org/10.1016/S0022-1694(03)00060-X

Wu, X., Ma, W., Xia, J., Bai, W., Jin, S. & Calabia, A. 2021 Spaceborne GNSS-R soil moisture retrieval: status, development opportunities, and challenges. *Remote Sensing* **13** (1), 45. Available from: www.app-ear.com.ar

Xu, C., Qu, J. J., Hao, X., Cosh, M. H., Prueger, J. H., Zhu, Z. & Gutenberg, L. 2018 Downscaling of surface soil moisture retrieval by combining MODIS/Landsat and in situ measurements. *Remote Sensing* **10**. https://doi.org/10.3390/rs10020210

Yang, Y., Zhu, J., Tong, X. & Wang, D. 2009 The spatial pattern characteristics of soil nutrients at the field scale. *Computer and Computing Technologies in Agriculture II*, Volume 1, pp. 125–134. https://doi.org/10.1007/978-1-4419-0209-2_14

Zare-Mehrjardi, M., Taghizadeh-Mehrjardi, R. & Akbarzadeh, A. 2010 Evaluation of geostatistical techniques for mapping spatial distribution of soil PH, salinity and plant cover affected by environmental factors in southern Iran. *Notulae Scientia Biologicae* **2**, 92–103. https://doi.org/10.15835/nsb244997

Ziadat, F. M. & Taimeh, A. Y. 2013 Effect of rainfall intensity, slope, land use and antecedent soil moisture on soil erosion in an arid environment. *Land Degradation & Development* **24**, 582–590. https://doi.org/10.1002/ldr.2239