

Article

GIS and Remote Sensing Aided Information for Soil Moisture Estimation: A Comparative Study of Interpolation Techniques

Prashant K. Srivastava ^{1,*}, Prem C. Pandey ², George P. Petropoulos ^{3,4},
Nektarios N. Kourgialas ^{5,*}, Varsha Pandey ¹ and Ujjwal Singh ¹

¹ Institute of Environment and Sustainable Development and DST-Mahamana Center for Excellence in Climate Change Research, Banaras Hindu University, Varanasi, Uttar Pradesh 221005, India; rschapterbook@gmail.com (V.P.); ujjwalrsmt@gmail.com (U.S.)

² Center for Environmental Sciences and Engineering, School of Natural Sciences, Shiv Nadar University, Greater Noida, Gautam Buddha Nagar, Uttar Pradesh 201314, India; prem26bit@gmail.com

³ Department of Soil & Water Resources, Institute of Industrial & Forage Crops, Hellenic Agricultural Organization, H.A.O. “Demeter” (former NAGREF), Directorate General of Agricultural Research, 1 Theofrastou St., 41335 Larisa, Greece; petropoulos.george@gmail.com

⁴ School of Mineral & Resources Engineering, Technical University of Crete, Kounoupidiana Campus, 73100 Chania, Crete, Greece

⁵ NAGREF-Hellenic Agricultural Organization (H.A.O.-DEMETER), Institute for Olive Tree Subtropical Crops and Viticulture, Water Recourses-Irrigation & Env. Geoinformatics Lab., 73100 Chania, Greece

* Correspondence: prashant.just@gmail.com (P.K.S.); kourgialas@nagref-cha.gr (N.N.K.)

Received: 21 February 2019; Accepted: 15 April 2019; Published: 18 April 2019



Abstract: Soil moisture represents a vital component of the ecosystem, sustaining life-supporting activities at micro and mega scales. It is a highly required parameter that may vary significantly both spatially and temporally. Due to this fact, its estimation is challenging and often hard to obtain especially over large, heterogeneous surfaces. This study aimed at comparing the performance of four widely used interpolation methods in estimating soil moisture using GPS-aided information and remote sensing. The Distance Weighting (IDW), Spline, Ordinary Kriging models and Kriging with External Drift (KED) interpolation techniques were employed to estimate soil moisture using 82 soil moisture field-measured values. Of those measurements, data from 54 soil moisture locations were used for calibration and the remaining data for validation purposes. The study area selected was Varanasi City, India covering an area of 1535 km². The soil moisture distribution results demonstrate the lowest RMSE (root mean square error, 8.69%) for KED, in comparison to the other approaches. For KED, the soil organic carbon information was incorporated as a secondary variable. The study results contribute towards efforts to overcome the issue of scarcity of soil moisture information at local and regional scales. It also provides an understandable method to generate and produce reliable spatial continuous datasets of this parameter, demonstrating the added value of geospatial analysis techniques for this purpose.

Keywords: spatial interpolation; geoinformation; mapping; monitoring soil moisture; soil water management; geographical information systems

1. Introduction

Soil moisture represents a vital component of the ecosystem sustaining life-supporting activities at micro and mega scales [1,2]. It is highly variable with spatial and temporal scales and depends upon the topographical, soil, land cover and climatic conditions [3,4]. Thus, soil moisture is an integral

part of the hydrological cycle and is essential for human and plant growth [5–7]. Measurement of this parameter is imperative to agricultural aspects due to its importance for early monitoring of drought warnings. Land use and land cover influence soil moisture spatiotemporal variability and can alter soil hydraulic properties because of changes in porosity and structure of the soil [8]. Soil moisture assessment in the root zone controlling the crop productivity provides information for deficits in soil moisture [9]. Hence, regular soil moisture monitoring provides appropriate irrigational facility to improve crop productivity and yield forecasting. Furthermore, soil moisture provides information, prediction and forecasting on flood before rainstorms. At the saturation level of soil, floods will likely to occur during the event of rain as its capacity to hold water has reached a limit and thus is unable to absorb an adequate amount of water [10–12]. Among others, Reference [13] stressed the importance of soil moisture information for meteorologists, since this parameter is related to weather changes, thereby providing a more accurate weather forecast. In addition, information on this parameter is important in biodiversity and ecosystems management [14].

These applications mentioned above show the importance of soil moisture at local, regional and global scale. Thus, precise, accurate and adequate knowledge on soil moisture at different scales is essential for agricultural crop productivity, flood monitoring and soil conditions/status. However, the lack of information on soil data at different scales for large regions is still an issue due to its extreme variability in both the spatial and temporal domain and at different observation scales. Soil moisture can be directly measured using wide range ground instruments [15]; yet, such approaches cannot fully describe the spatial and temporal variability of soil moisture at larger scales. On the other hand, remote sensing techniques have the advantage of achieving simultaneously satisfactory sampling frequency and global coverage [7,16,17]. Nonetheless, most of the relevant operational products available today are providing soil moisture at coarse resolution. To acquire high spatial resolution soil moisture at regional and local scales, spatial interpolation and GIS can play an important role. Spatial interpolation and GIS emerged as powerful tools in remote sensing technology demonstrating their capability to provide spatial continuous data required for environmental monitoring [18].

For estimation of attribute values at unsampled location, spatial interpolation of in-situ measurements from sampled locations is necessary to achieve continuous spatial data. This necessity is added when the discrete surface has a different spatial resolution, cell size or orientation as required, when the continuous surface is represented by a data model other than required and when the available datasets do not cover the regions of interest completely [19,20]. In these instances, spatial interpolation methods act as a solution and provide continuous data at the unsampled sites (unknown locations) by estimating environmental variables from point observations within the same region [20]. Several spatial interpolation as well as extrapolation methods were developed and applied in different research disciplines including geo-statistics [21]. There is a variety of such methods which are proposed as more suitable in each case according to the specific data types or specific variables to be interpolated. Thus, is difficult to assign spatial interpolation methods as the “best practices” based on a given datasets without a prior investigation of their accuracy [22]. However, spatial interpolation has a profound impact on several factors including sample size, sample design and the nature of the data. Moreover, the predictive performance of the method were not consistent as shown in previous studies, while as the variation increases, the accuracy of all methods changes either increases or decreases [23]. Yet, there is no proved consistent findings about variation and how the previously mentioned factors affect the performance of the spatial interpolation methods. Therefore, it is difficult to select an appropriate spatial interpolation method for a given specific input dataset due to variation within the data [20].

There have been some research studies focusing on interpolating soil moisture [22,24–27]. For example, [22] demonstrated that ordinary kriging and Inverse Distance Weighting (IDW) methods employed for the estimation of soil moisture at complex terrain results in poor performance due to low spatial autocorrelation of soil moisture at small catchment scale [24,25,28]. Another study [27], compared different interpolation methods (inverse distance weighting, multifarious forms of kriging, regularized spline with tension and thin plate spline) for estimating soil moisture in an area with

complex topography in southwest China. Their results indicated that inverse distance weighting had the best performance, at least this was the case in their study.

The present study builds on studies such as the above, aiming at developing a geo-spatial database for soil moisture at district level and assessing a range of widely used spatial interpolation techniques for soil moisture estimation using GPS-aided information. Four spatial interpolation methods, namely Inverse Distance Weighting (IDW), Spline, Ordinary Kriging models and Kriging with External Drift (KED) are compared in estimating soil moisture. As a case study, Varanasi city located in India is used, and for this region to our knowledge for the first time a soil moisture destitution map is developed for the area taking into account all the particular environmental characteristics of the region. An additional added value of our study is also the use of remote sensing techniques to define Digital Elevation Model (DEM) and Land Cover maps and the incorporation of soil organic carbon information as a secondary variable.

2. Materials and Methods

2.1. Study Area

The study area chosen is Varanasi City, India. The study area extends from 25°10'3" and 25°35'1" N to 82°40'5" and 83°12'1" E. It covers a total area of 1535.00 km² which 1371.22 km² is under rural land and 163.78 km² is under urban land and it is elevated above mean sea level at 80.71 m (Figure 1). The River Ganga here flows south to North, having the highest flood level at 73.90 m (1978) and the lowest river water level is approximately 58 m. It is a part of Indo-Gangetic plain covering parts of alluvial deposit of river Ganga and Varuna makes the favorable conditions like fertile soil, plain topography for settlement, availability of fresh water have intense bearing on the high population density along with the historical development of the town as an oldest pilgrimage city in India. It consists mainly of sand, silt and clay interspersed by pellets stones at a few places nearby rivers. Climatically the region has sub-tropical monsoonal climate characterized by seasonal extremities. It has humid sub-tropical climate and experiences large variation between summer and winter temperatures. The total annual precipitation is 10,582 mm and humidity is ~70% and mean maximum annual temperature is ~32 °C and mean minimum annual temperature is ~19 °C. The studied area is situated on the crescent shape bank of perennial river Ganga up to a stretch of 2525 km² which has high seasonal variations in discharge due to monsoonal impact.

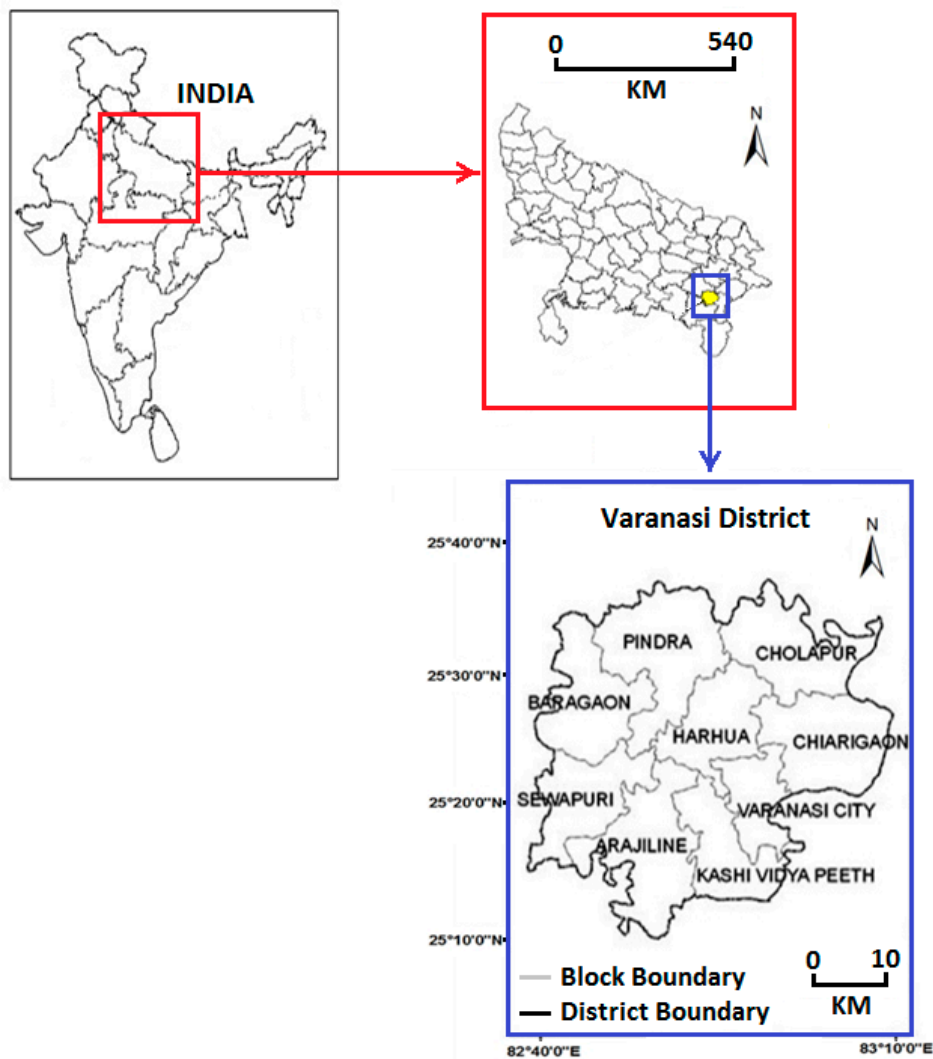


Figure 1. Location map of the study area.

2.2. Datasets and Methodology

The entire Varanasi district, including nine district blocks, namely, Varanasi City, Arajilina, Sewapuri, Baragaon, Pindra, Chhapra, Harhua, Chharaigaon and Kashi Vidya Peeth (KVP) was surveyed to collect soil samples with 5 cm diameter core pipe and up to the depth of 5 cm and 20 cm from the surface. From each block, 10 locations were chosen randomly for consideration with a horizontal distance of 5 to 9 km among locations and in different directions. The Varanasi city block was excluded from sample collection as it is a fully urban area. Major soil samples were taken from paddy fields and some sample points from barren land and plantation area because this study has its main focus on soil moisture and its related attributes so non-arable lands are of less importance. The entire fieldwork was completed during the months of August and September 2015 for a total of 82 locations (Figure 2). Out of the total 82 locations, 54 locations data were used for calibration and 28 locations data was used for validation following the thumb rule of 2/3 for calibration and 1/3 for validation. Soil moisture (in % of vol. vol.⁻¹) data on field for every location was measured by Stevens Hydraprobe at 5 cm and 20 cm depths. GPS (global positioning system) was used to record coordinates of every location that was used further in map processing in GIS. The data processing and analysis was worked with ArcGIS v10.1 (ESRI, Redlands, CA, USA) and ENVI v5.2 (Exelis, Boulder, CO, USA). Statistical analysis was performed using spatial analyst of ArcGIS v10.1, RStudio platform and Microsoft Excel packages.

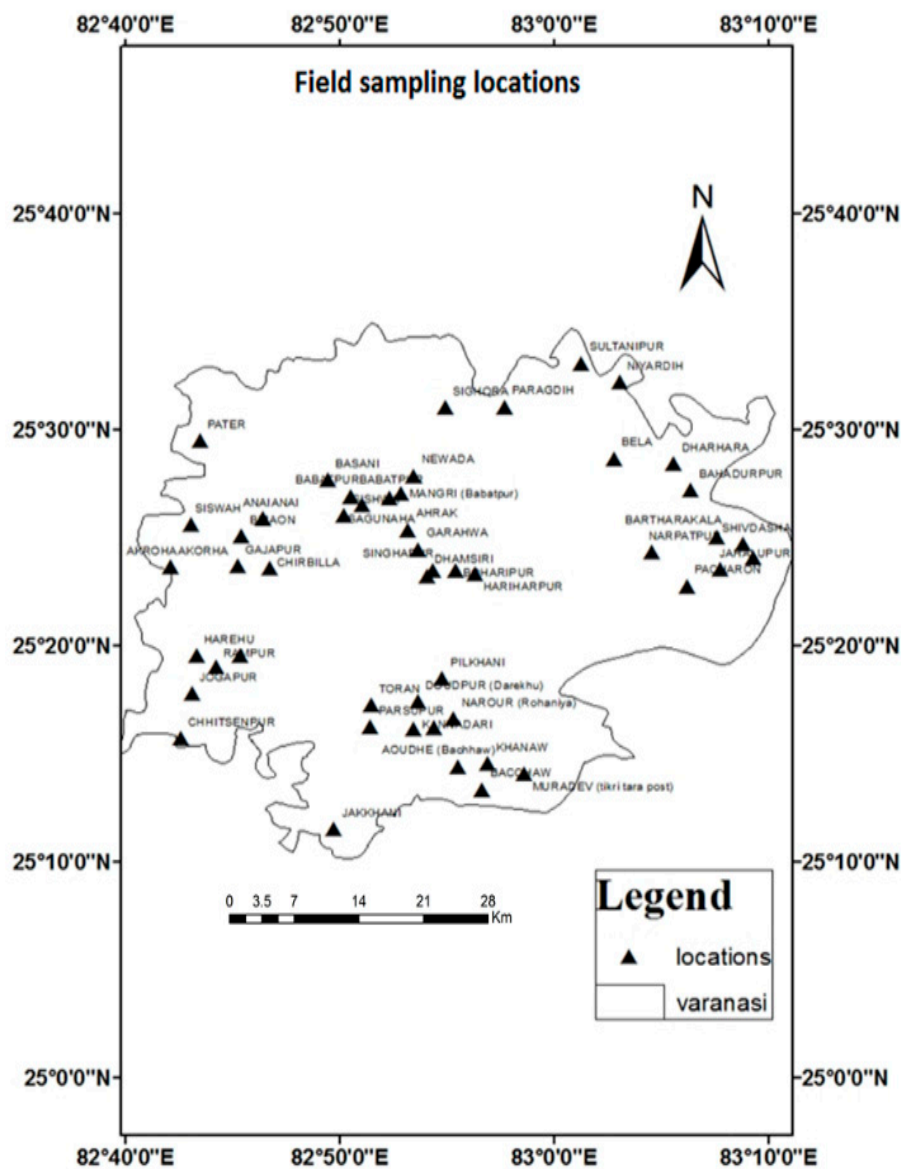


Figure 2. Sampling site locations in the study area.

Soil moisture was estimated using mean values from soil moisture measurements that took place at the above mentioned 82 locations of the study area. The soil moisture measurements took place daily during the months of August and September of 2015. Also, in the studied interpolation techniques, for each location, the mean values of soil moisture at the two different depths (5 and 20 cm) were used as input.

In addition, the above mentioned 82 soil sampling locations were used to obtain the appropriate information regarding the soil organic carbon. In turn, soil organic carbon information was incorporated in KED interpolation method as a secondary variable. Also, in this study a series of statistical metrics (Percent bias (%Bias), root mean square error (RMSE), mean absolute error (MAE)) were utilized in order to determine the performance of the different interpolation models.

2.3. Methodology

2.3.1. Soil Moisture and Organic Carbon Estimation

In this study, soil moisture measurements were conducted using the Stevens hydroprobe. This provides a measurement of the average volumetric water content along the length of the

waveguide. Among the most important advantages of this technique are that it is non-destructive to the study site and is not labour intensive [29,30].

The amount of organic carbon present in the soil samples was determined applying the Walkley-Black chromic acid wet oxidation method. The oxidisable organic matter in the soil is oxidized by 1N K₂Cr₂O₇ solution to estimate soil organic carbon. For preparation of 1 N (normality) Potassium dichromate solution, 49.040 g of K₂Cr₂O₇ was dissolved in the deionized water and then the volume of the volumetric flask was made to 1 litre. Second solution was prepared for 0.4 N ferrous ammonium sulphate. 0.4 N ferrous ammonium sulphate solution, 112 g of FeSO₄ (NH₄)₆·7H₂O was dissolved in 800 mL of deionized water. Then, 15 mL of concentrated H₂SO₄ was added to facilitate proper dissolution. Following this procedure, the volume of the solution was made to 1 litre and thereafter, the solution was stored in a dark bottle.

The soil sample was crushed and filtered using a 0.42 mm sieve followed by drying in the oven. Thereafter, dried weighted soil sample was transferred to a 250 mL Erlenmeyer conical flask. 10 mL of potassium dichromate was added to the dried soil followed by gentle swirling to allow proper mixing of the solution with the soil. Then, 20 mL of concentrated H₂SO₄ was added to the soil solution. Following this it was allowed to stand for 1 hour for its proper digestion. Then 200 mL of distilled water was added and allowed to cool. It was then mixed with 2 drops of diphenyl indicator and titrated against 0.4 M ferrous ammonium sulphate. Before analysing the samples, a blank was tested and titrated against the same. The amount of organic carbon in the soil samples was calculated using the following formula (Walkley-Black chromic acid wet oxidation method):

$$\text{Organic Carbon (\%)} = \frac{0.003 \text{ g} \times N \times 10 \text{ mL} \times \left(1 - \frac{T}{S}\right) \times 100}{\text{ODW}} \quad (1)$$

where; N is the normality of K₂Cr₂O₇ solution, T is the volume of FeSO₄ used in sample titration (mL), S is the volume of FeSO₄ used in blank titration (mL) and ODW is the oven-dried sample weight.

The soil organic carbon was estimated using mean values from soil organic carbon analysis-measurements that took place at the above mentioned 82 soil sampling locations of the study area (complex samples at two different soil depths 5 cm and 20 cm for each location). The sampling campaign took place during the months of August and September of 2015. In turn, the mean values of the soil organic carbon for each of the 82 locations were used as a secondary variable in KED interpolation method.

2.3.2. Digital Elevation Model (DEM)

A DEM helps in the demonstration of the topographic data and plays an important role in the modelling and prediction of floods [31]. DEM involves data in a geodetic coordinate system that is in longitude and latitude. The Shuttle Radar Topography Mission (SRTM) digital elevation data, produced by NASA (The National Aeronautics and Space Administration) originally, is a major breakthrough in digital mapping of the world, providing a major advance in the accessibility of high quality elevation data for large portions of the tropics and other areas of the developing world. The SRTM 90 m DEM provided in mosaicked 5° × 5° tiles was used in this study. The vertical error of the DEM's is reported to be less than 16 m. In our study, the DEM of SRTM was downloaded from earth explorer official website which is easily available online after minimum formalities.

2.3.3. Database for Land Use Land Cover

The Landsat images were downloaded from the United States Geological Survey (USGS) portal (<http://www.usgs.gov/>). Landsat 8 (L8) data was incorporated with the study of image classification with total eight bands. L8 was launched in 2013 and it has Enhanced Thematic Mapper (ETM+) sensor, 185 km swath, 8 bits, SWIR 30 m and TIR 60 m. World Geodetic System 1984 (WGS84) was used for geocentric reference (Table 1).

Table 1. Technical details of Landsat 8 dataset used in this study.

Sensor Type	Path	Row	Spatial Resolution (in Meter)	Date of Acquisition	Season	Time
Landsat 8	142	39	30	22 April 2013	summer	23:31

2.3.4. Image Classification

Following the data acquisition, bands stacking and image subset was applied to the images to facilitate computational efficiency in data handling. After this step, land use/cover classes to be included in the classification maps were decided and representative training sites for each of those classes were selected. Approximately 150 pixels per class (600 total pixels) were selected from homogeneous areas. The separability of the selected training points for all cover classes were examined in ENVI (v 5.2, Exelis, Boulder, CO, USA) software. Then, a pixel based supervised image classification that is, Support Vector Machine with Radial Basis Function (SVM RBF) [32,33] was implemented to the acquired Landsat images in ENVI using the selected training data. SVMs can produce accurate and robust classification results on a sound theoretical basis, even when input data are non-monotone and non-linearly separable. So, they can help to evaluate more relevant information in a convenient way. Since they linearize data on an implicit basis by means of kernel transformation, the accuracy of the results does not rely on the quality of human expertise judgment for the optimal choice of the linearization function of non-linear input data. SVMs operate locally, so they are able to reflect in their score the features of single companies, comparing their input variables with the ones of companies in the training sample showing similar constellations of financial ratios. Although SVMs do not deliver a parametric score function, its local linear approximation can offer an important support in recognising the mechanisms linking different financial ratios with the final score of a company. For those reasons SVMs are regarded as a useful tool for effectively complementing the information gained from classical linear classification techniques [34].

2.4. Spatial Interpolation Methods

There are several methods developed for spatial interpolation in various disciplines with a number of different terminologies used to distinguish them, including “interpolating” and “non-interpolating” methods or “interpolators” and “non-interpolators” [35]. The spatial interpolation methods covered in this study focuses on four methods namely IDW, spline, kriging and linear spatial interpolation methods. All of these methods are widely used in interpolation studies and their implementation was incorporated in most relevant software packages.

Spatial interpolation methods are widely used in soil analysis in environmental studies, briefly fall into three categories: 1) non-geostatistical methods, 2) geostatistical methods and 3) combined methods. In geostatistics, the methods that are capable of using secondary information are often referred to as “multivariate,” while the methods that do not use the secondary information are called “univariate” methods. Here, it must be noted that multivariate usually refers to more than one response variable, despite of the fact that in some references it also refers to more than one explanatory variable (usually referred to as multiple variables).

2.4.1. Inverse Distance Weighting (IDW)

The inverse distance weighting or inverse distance weighted (IDW) method estimates the values of an attribute at unsampled points using a linear combination of values at sampled points and weighted by an inverse function of the distance from the point of observation to the sampled points. The assumption is that sampled points closer to the unsampled point are more similar to it than those further away in their values. The weights can be expressed as:

$$\gamma_i = \frac{1/d_1^p}{\sum_{i=1}^n \frac{1}{d_1^p}} \quad (2)$$

where, d_i is the distance between x_0 and x_i , p is a power parameter and n represents the number of sample points used for the estimation.

The main factor affecting the accuracy of IDW is the value of the power parameter [36]. Weights diminish as the distance increases, especially when the value of the power parameter increases, so nearby samples have a heavier weight and have more influence on the estimation and the resultant spatial interpolation is local [37]. The choice of power parameter and neighbourhood size is arbitrary [38]. The most popular choice of p is 2 and the resulting method is often called inverse square distance or inverse distance squared (MacEachren and Davidson). The power parameter can also be chosen based on the error measurement (e.g., minimum mean absolute error, resulting the optimal IDW) [18]. The smoothness of the estimated surface increases as the power parameter increases and it was found that the estimated results become less satisfactory when p is 1 and 2 compared with p is 4 (Ripley, 1981). IDW is referred to as “moving average” when p is zero [36], “linear interpolation” when p is 1 and “weighted moving average” when p is not equal to 1 [21]. The IDW works well with regularly spaced data but it is unable to account for clustering [36].

2.4.2. Splines and Local Trend Surfaces (SPLINE)

The splines consist of polynomials with each polynomial of degree p being local rather than global. The polynomials describe pieces of a line or surface (i.e., they are fitted to a small number of data points exactly) and are fitted together so that they join smoothly [20,38]. The places where the pieces join are called knots. The choice of knots is arbitrary and may have a dramatic impact on the estimation [20]. For degree $p = 1, 2$ or 3 , a spline is called linear, quadratic or cubic respectively. Typically the splines are of degree 3 (i.e., are cubic splines) [38]. The local trend surfaces (LTS) fit a polynomial surface for each predicted point using the nearby samples. Splines are deterministic with locally stochastic properties. Splines are piece-wise functions using a few points at a time. The interpolation predictions can be quickly calculated and predictions are very close to the values being interpolated, providing the measurement errors associated with the data are small [20]. Splines retain small-scale features but there are no direct estimates of the errors. Spline implementation to data on a grid requires special care because if the dataset does not have the direct information needed for reliable prediction the dataset cannot provide direct information on residual variance [36]. Exact splines may produce local artefacts of excessively high or low values. These artefacts can be removed using the Thin Plate Splines (TPS), where an exact spline surface is replaced by a locally smoothed average [20]. TPS can also be extended to include multivariate spline function [39]. TPS may provide a view of reality that is unrealistically smooth and thus generate misleading results [20]. Spline technique provides enough flexibility for local geometry analysis that can often be used as input to various process-based models [40]. However, most of the surfaces or volumes are neither stochastic nor elastic media but are the result of a host of natural (e.g., fluxes, diffusion) or socioeconomic processes. Improvements in accuracy and realism can be expected by employing spatially-variable adaptive interpolation and by further developments in model-based interpolation [39].

2.4.3. Kriging: A Geo-Statistical Interpolators

Geostatistical approaches are used to: 1) describe spatial patterns and interpolate the values of the primary variable at unsampled locations; and 2) model the uncertainty or error of the estimated surface. Originally in geostatistics, Kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modelled by a Gaussian process governed by prior covariances to optimize smoothness of the fitted values. Under suitable assumptions on the priors, Kriging gives the best linear unbiased prediction of the intermediate values. Interpolating methods based on other

criteria such as smoothness need not yield the most likely intermediate values. The method is widely used in the domain of spatial analysis. The technique is also known as Wiener–Kolmogorov prediction, after Norbert Wiener and Andrey Kolmogorov.

All kriging estimators are variants of the basic following equation:

$$z(x_0) - \mu = \sum_{i=1}^n \lambda_i [z(x_i) - \mu(x_0)] \quad (3)$$

where: μ is a known stationary mean, assumed to be constant over the whole domain and calculated as the average of the data (Wackernagel, 2003). The parameter λ_i is kriging weight; n is the number of sampled points used to make the estimation and depends on the size of the search window; and $\mu(x_0)$ is the mean of samples within the search window.

2.4.4. Ordinary Kriging (OK) and Kriging with an External Drift (KED)

Ordinary kriging (OK) is a linear interpolator that estimates a value at a point of a region for which the variogram is known, without prior knowledge about the mean of the distribution. In this method a random function model is used, in which the bias and error variance can both be calculated and then weights are chosen for the nearby samples such that they ensure that the average error for the model is zero and the modelled variance is minimized. In order for the estimator to be unbiased in this technique implementation, the sum of these weights needs to equal one [35,41].

The kriging with an external drift (KED) incorporates the local trend within the neighbourhood search window as a linear function of a smoothly varying secondary variable instead of as a function of the spatial coordinates [42]. The trend of the primary variable must be linearly related to that of the secondary variable. This secondary variable should vary smoothly in space and is measured at all primary data points and at all points being estimated. Kriging has few advantages and also some drawbacks. It provides the best linear unbiased estimate. It also provides a measure of the error or uncertainty at the unsampled points. However, it assumes stationarity of data, which is usually not true, although this assumption can be relaxed with specific forms of kriging.

Generally, Kriging requires a large number of samples, at least 100, to produce a reliable estimation of variogram [38]. This limitation could be overcome by using in KED the Residual Maximum likelihood (REML) variogram because predictions based on REML variograms are generally more accurate than those from the conventional moment variograms with fewer than 100 samples [43]. For spatial prediction of soil variables where the local mean can be expressed a linear function of some auxiliary (external drift) variable, the state-of-the-art is to estimate the spatial covariance parameters of the residual variation by residual maximum likelihood (REML), as it is described that is less sensitive to outliers [44]. In such cases, a sample size of 50 appears adequate [43]. Thus, in this study the REML estimator in KED interpolation method was used. In KED the secondary information provides information only about the primary trend of the point of interest and tends to strongly influence the estimate especially when the estimated slope of the local trend model is large. In the present study the secondary information that was used concerns the soil organic carbon.

2.5. Statistical Criteria

In the following section a short description of the different statistical metrics used in our study, namely the %Bias, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) is provided. Percent bias (%Bias or PBIAS) measures the deviation of the simulated values from the observed ones [45]. The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate overestimation bias, whereas negative values indicate model underestimation between simulated and observed. The result is given in percentage (%) [43].

RMSE is one of the measures used to assess accuracy of spatial analysis and remote sensing products [45]. The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a widely

used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed [46]. RMSE represents the standard deviation of sample and the differences between predicted values and observed values [46]. These individual differences are called residuals when the calculations are performed over the data sample that was used for estimation and are called prediction errors when computed out-of-sample [45]. RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power [47]. RMSE is a good measure of accuracy but only to compare forecasting errors of different models for a particular variable and not between variables, as it is scale-dependent [48]. The lower the RMSE value, the better performance of the model is [49]. Last but not least, MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction and measures accuracy for continuous variables [50]. It is a quantity used to measure how close forecasts or predictions are to the eventual outcomes and at the same time is a common measure of forecast error of a model. It is a linear score which means that all the individual differences are weighted equally in the average [45].

3. Results

This section presents the main study results. All the models and techniques utilized in this study are evaluated along with map interpretation of DEM, LULC, soil moisture and organic carbon. This is followed by the statistical analysis performance to get a conclusion for the spatial interpolation techniques. Digital elevation map of Varanasi district is showing clearly that the average elevation varies from 47 m to 99 m approximately across the region (as shown in Figure 3). In general, elevation has declining trend from east to west due to its geological characteristics and topographical feature of the Vindhayan plateau formation at the west. On an average, south western and north western parts have more elevation above mean sea level then the rest of the areas. It is well known that the Eastern part has the holy river Ganga flowing from north east to south west in this region. Therefore, it is likely to have lower elevation due to lateral erosional work of the river covering southern parts of the study site.

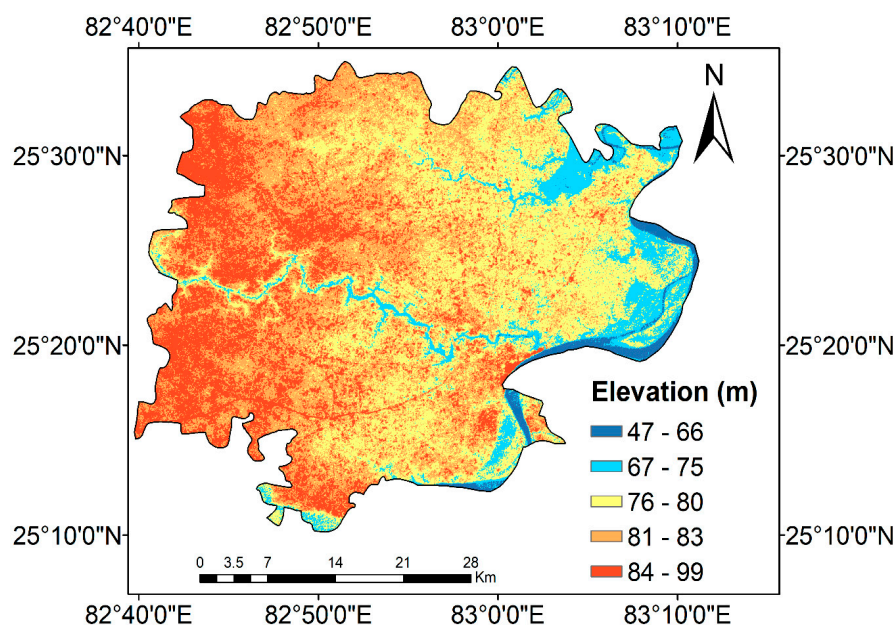


Figure 3. Digital elevation map of the study area.

3.1. Nature of Land Use and Land Cover of the Varanasi

The land use land cover (LULC), (2013) map of Varanasi district (Figure 4) shows that the major area is dominated by agriculture classes. Generally open space, fallow-land and other vegetative covers were taken as agriculture only so that generalization as per the study objectives can be done (see Table 2).

Some vegetation cover that is mainly plantation, denoted with a green colour was spread inside and nearby city area especially. Major urban agglomeration can be seen around Varanasi city block and northern Easter part of Kashi Vidya Peeth block along the stretches of river Ganga. While remaining blocks of Harhua, Pindra, Cholapur, Arajiline, Baragaon, Chiraigaon and Sewapuri have fewer urban parts except few central market locations. Overall nature of urbanization is considered spurious rather than planned. The relation between mean SM and Land use land cover showed that marginally highest variation of SM is found over urban area followed by agriculture, while least variation is detected over sandy areas.

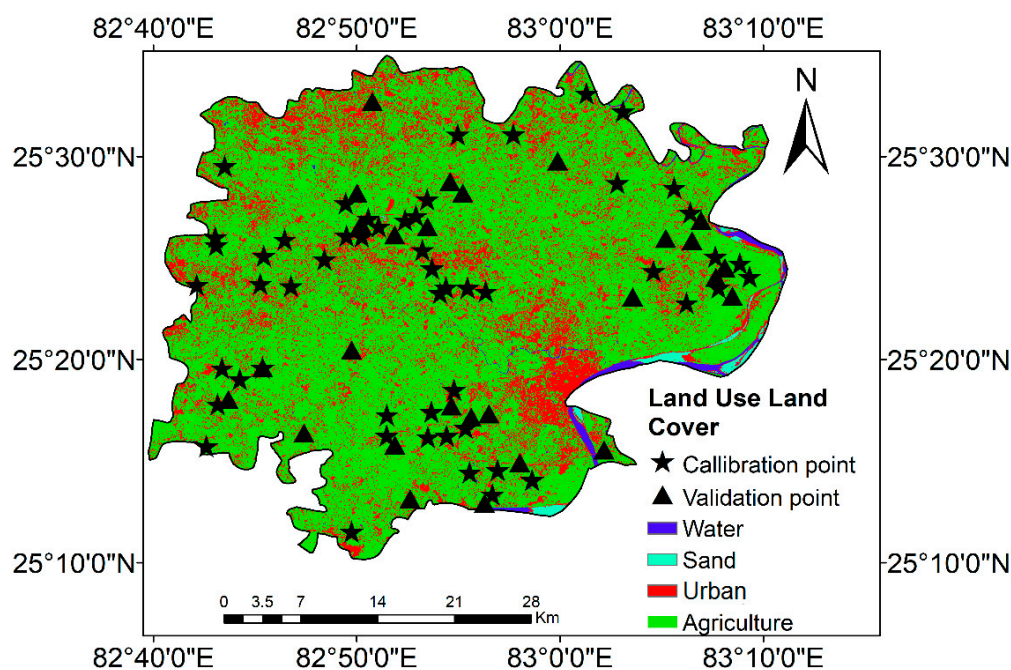


Figure 4. LULC map of the study area with calibration and validation dataset.

Table 2. Land Use Land Cover Statistics with mean SM (%) distribution.

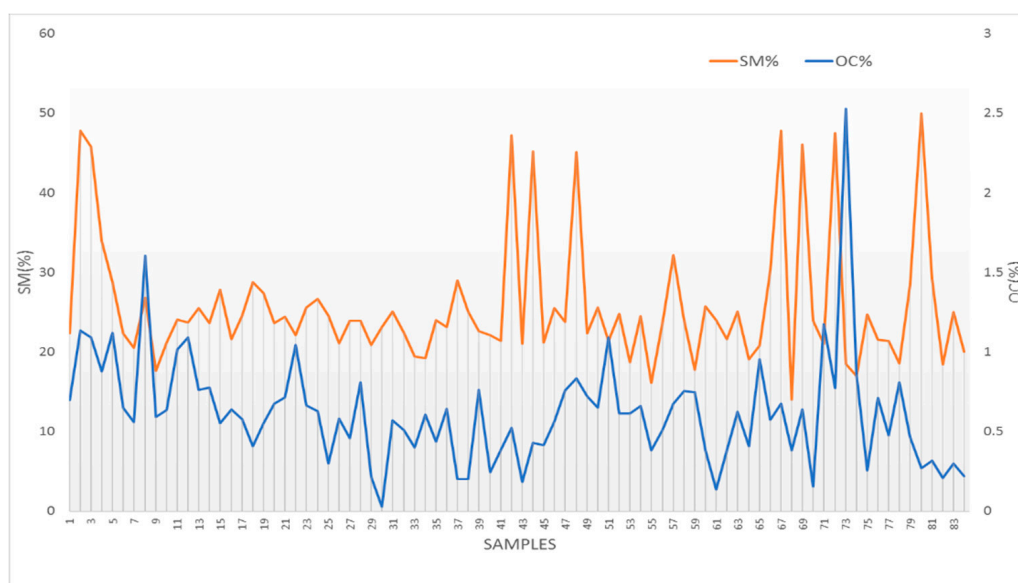
Serial No.	Cover Class	Area in Hectares	SM _{Min.}	SM _{Max.}	SM _{Mean}
1	Urban	116,823.6	15.52	46.67	25.56
2	Water	22,478.4	–	–	–
3	Sand	3474.99	23.53	35.15	30.33
4	Agriculture	52,2763.4	15.60	46.34	25.43

3.2. Soil Moisture Distribution across Study Area

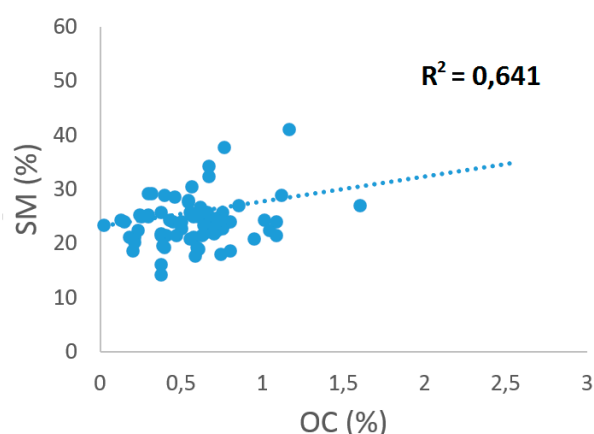
3.2.1. Soil Organic Carbon Distribution the Secondary Variable of KED Method

Regarding the kriging with an external drift (KED), as mentioned in the methodology section, the trend of the primary variable must be related to the secondary variable (soil organic carbon). Distribution of soil organic carbon is varying from 0.03% to 2.5% across the study region showing fluctuation over space. It is interesting to note that there are some areas like parts of Baragaon, Pindra and KVP, where soil moisture is high, organic carbon is also high where as some areas like Varanasi city block and Chiraigaon, is showing contrary pattern where soil moisture is high but organic carbon is low. Possible reason could be the use of ground water for various uses in city area lead to low soil moisture. Blocks such as Harhua, Sewapuri, western-central Baragaon and northern Cholapur have relatively low soil moisture as well as low organic carbon that could have possibly due to its agricultural dominance.

Figure 5a illustrates the soil moisture trends (mean values of soil moisture at the two different depths) and soil organic carbon values at the 82 selected soil moisture field-measured locations. As shown, soil moisture and organic carbon is following a similar trend of ups and downs except at few locations of Ahrak, Bahrapur, Sonbarsha, Siswa, Paragdih, Sulatnipur village. Figure 5b illustrates the correlation between soil moisture and soil organic carbon values. The above, indicates the appropriateness of using soil organic carbon as a secondary variable in KED method. Some mismatching between OC and SM could be ascribed to the agricultural practices in the area such as tillage and irrigation of the crops. Due to overturning of the soil in some areas because of tillage practices some poor agreement found between the two variables used in this study. Further in some areas, due to irrigation some poor relation is found between OC and SM can be also seen in the graph. Figure 5b provided a scatterplot of SM and OC, which indicates that some relationship exist between the two parameters and hence can be used as an external variable for soil moisture interpolation in KED method.



(a)



(b)

Figure 5. (a) Relationships between soil moisture (SM %) and organic carbon content (OC %), at the 82 soil moisture field-measured locations/samples; (b) Scatterplot between soil moisture (SM) and organic carbon content (OC).

From all the maps of the soil moisture produced from all four models of spatial interpolations, of both the data set (calibration and validation), it is evident that soil moisture is decreasing from east to west while. Looking at the overall map of soil moisture including all sample points (Figure 6) is showing that the eastern stretch from north to south covering Varanasi city block, parts of KVP, eastern Chiraigaon, central Sewapuri, northern-eastern Baragaon block have high soil moisture ranging from 25% to 50%.

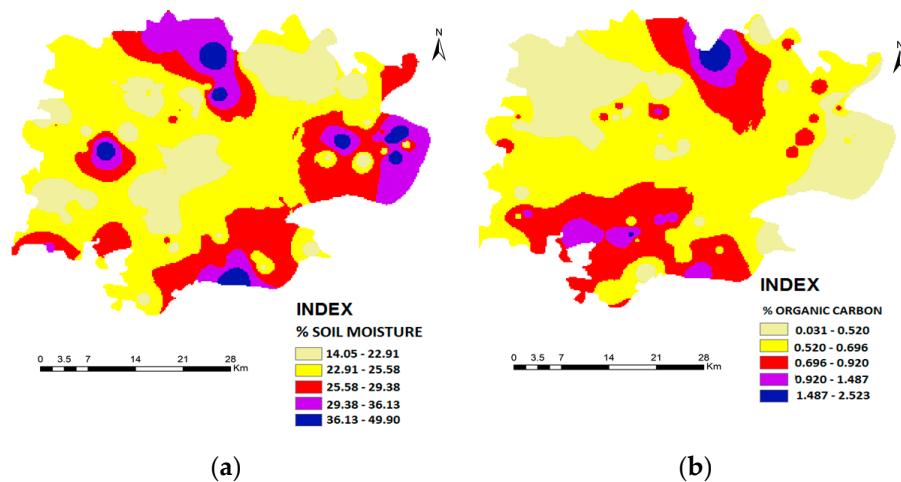


Figure 6. (a) Soil moisture and (b) Organic carbon distribution map of the study area, using the KED interpolation method.

It is evident from the field visit conducted at the site that these are mainly urban in nature and dominated by built up areas. This can also be clearly seen from the LULC map (Figure 4). Thus, it can be inferred that area of high urban agglomerations have a higher amount of soil moisture in comparison to the agricultural lands as plants utilize SM for their growth. Lands near to the water bodies and river beds, are also in line with the same trend of high soil moisture. From the DEM map (Figure 3) of Varanasi district, it is visualized that areas with a high altitude have a negative bearing on soil moisture and vice versa.

3.2.2. Performance Evaluation of Spatial Interpolation Techniques

By visualization of the four soil moisture interpolation maps (IDW, OK, KED and Spline) as shown in Figure 7a–d, it can be concluded that KED and OK has better prediction than IDW and Spline. KED and OK have both captured more regional variation of soil moisture whereas the splines technique has smoothed the map leading to rough estimation and very generic profile of soil moisture map. IDW results capture more variation of local soil moisture in comparison to Splines method. Figure 7a–d represents the soil moisture maps of cross calibration data using four different techniques of spatial interpolation, namely IDW, OK, KED and Spline.

In this study, the accuracy and robustness of all the four spatial interpolation methods of soil moisture was evaluated. The studied spatial interpolation results were discussed in terms of their assumptions and applicability using statistical performance indices such as RMSE and %BIAS. Results indicated a better performance of KED in comparison to IDW, OK and Spline, at least in our study. The final model variogram for OK and KED was chosen on the basis of the lowest RMSE from cross validation. Regarding the variogram characteristics of OK and KED the distance after which data are no longer correlated (range) was equal to 1700 and 1900 m, respectively. Moreover, the RMSE value is highest (9.32) for IDW followed by Spline, OK and lowest for KED (8.69). In terms of RMSE, IDW performs the least in interpolating soil moisture. In terms of %BIAS, it is evident, at least in our study, that Spline is more sensitive to the significant over prediction than the other three methods (see Table 3). The analysis of the distribution of RMSE and %BIAS shows that the soil moisture was

effectively predicted with KED techniques with low bias and RMSE. IDW, OK and KED performs underestimations of model error as their values are in negatives. The MAE statistics shows that KED performed better than OK, IDW and Spline as their values are more than the KED. Variance (difference between RMSE and MAE) of the model was high in KED and OK (both have values of 3.17 percentage) followed by IDW (1.18). Therefore, it can be said that the prediction of soil moisture values was better using the KED method than with the rest of the techniques.

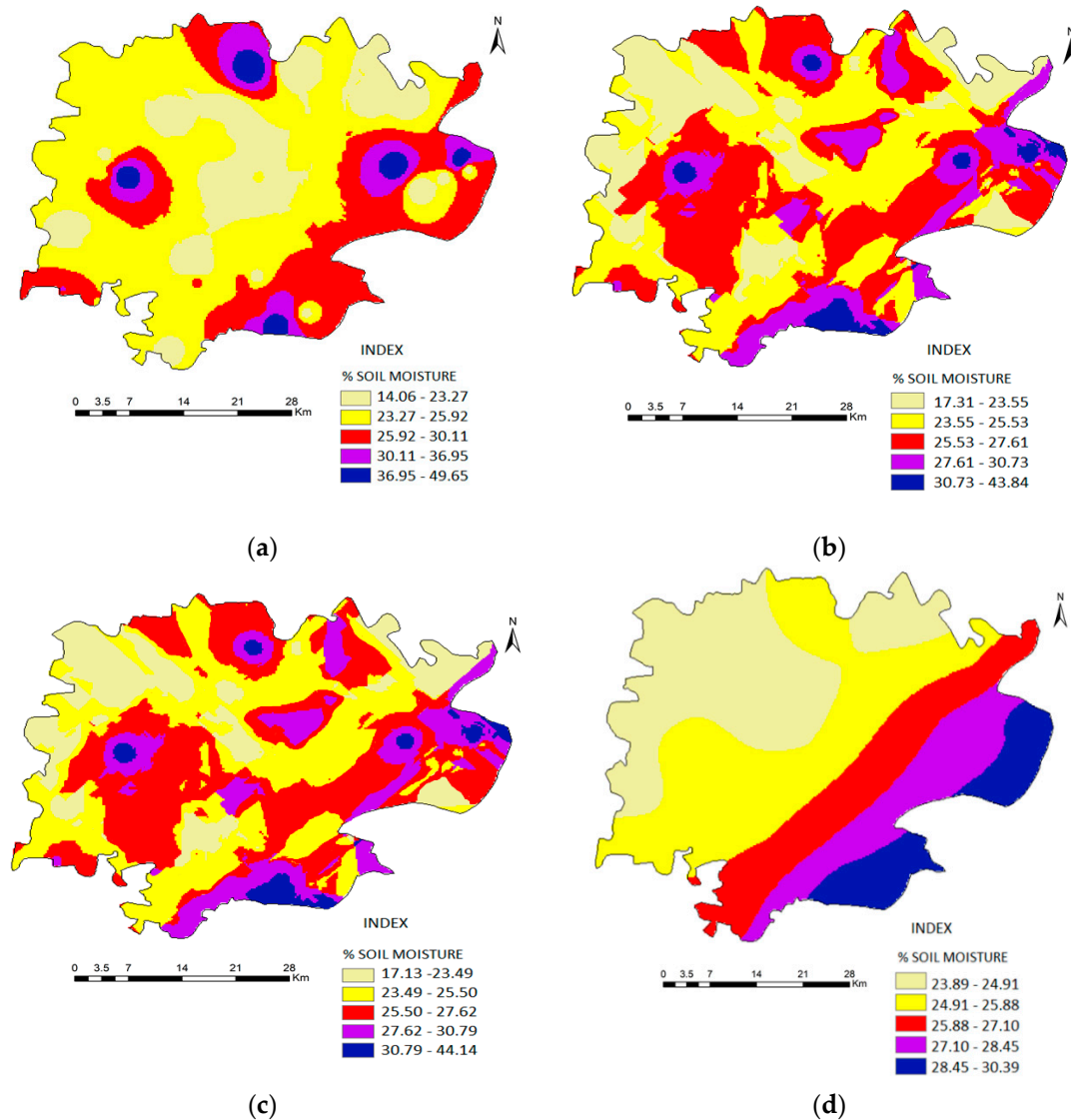


Figure 7. Soil moisture distribution using different spatial interpolation methods (a) IDW, (b) Ordinary Kriging, (c) Kriging with external drift (KED) and (d) Spline method.

Table 3. Performance statistics of the spatial interpolation methods.

	IDW	SPLINE	OK	KED
RMSE	9.32	8.82	8.73	8.69
%BIAS	-0.3	0.9	-0.9	-0.9
MAE	6.14	8.82	5.56	5.52

4. Conclusions

This study clearly demonstrated that the investigation of soil moisture distribution using GIS spatial interpolation techniques is an integral part of understanding the relationships between soil characteristics. Soil moisture is important not only to agriculture but also for hydrology and climatology. Remote sensing devices, such as satellite radiometers, are useful tools to obtain soil moisture information over a large region. However, effective ground truth calibrated data for satellite are lacking and this study fills this gap in providing values for soil moisture at local spatial level. Soil moisture distribution over the study area shows that higher elevation has lower soil moisture content and vice versa and its spatial distribution is not uniform over space and time. Even organic carbon profile has not certain pattern of distribution but in general, it is low in areas of vegetation cover and high in urban regions. The performance of spatial interpolation techniques depends on several parameters such as sample size, sampling design, sample spatial distribution, data (density and variation), surface type, data quality, input data uncertainty, poor choice or implementation techniques. Further data quality depends upon the distribution, accuracy, variance and range and spatial correlation for its performance. In this study, not all these mentioned parameters were incorporated individually due to constraints.

The study of spatial interpolation of soil moisture based on ground measurement data of soil moisture revealed that kriging with external (KED) is performing better than ordinary kriging (OK), inverse distance weighting (IDW) and Spline methods in terms of model accuracy and performance using cross validation techniques. KED is performing well over OK when secondary variable of organic carbon was added. This study can be extended in future, incorporating many parameters like soil texture, NDVI (Normalized Difference Vegetation Index), LST (Land Surface Temperature), bulk density and can also be calibrated by satellite data of soil moisture. This study can also lead a way for spatial analysts to see which method can perform well while adding more parameters and what will be the threshold.

Author Contributions: Conceptualization, P.K.S.; methodology, P.K.S. and G.P.P.; software, P.K.S., V.P.; validation, P.K.S., P.C.P., V.P. and U.S.; formal analysis, P.K.S. and G.P.P.; investigation, P.K.S., N.N.K., G.P.P.; resources, P.K.S.; data curation, P.C.P., V.P.; writing—original draft preparation, N.N.K., G.P.P., P.C.P. and P.K.S.; writing—review and editing, N.N.K. and G.P.P.; visualization, P.K.S.; supervision, P.K.S.; project administration, P.K.S. and G.P.P.; funding acquisition, P.K.S.

Funding: PKS contribution was supported by RESPOND, ISRO (Project code M-21-172). GPP's contribution was supported by the FP7- People project ENViSiON-EO "Enhancing our Understanding of Earth's Land Surface InteractiONs at Multiple Scales Utilising Earth Observation" (project reference number 752094).

Acknowledgments: We are grateful to the anonymous reviewers for their feedback that helped improving the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Falkenmark, M.; Rockström, J. *Balancing Water for Humans and Nature: The New Approach in Ecohydrology*; Earthscan: London, UK, 2004. [[CrossRef](#)]
2. Bao, Y.; Lin, L.; Wu, S.; Deng, K.A.K.; Petropoulos, G.P. Surface soil moisture retrievals over partially vegetated areas from the synergy of sentinel-1 & landsat 8 data using a modified water-cloud model. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 76–85. [[CrossRef](#)]
3. Western, A.W.; Zhou, S.-L.; Grayson, R.B.; McMahon, T.A.; Blöschl, G.; Wilson, D.J. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.* **2004**, *286*, 113–134. [[CrossRef](#)]
4. Piles, M.; Petropoulos, G.P.; Sanchez, N.; González-Zamora, A.; Ireland, G. Towards improved spatio-temporal resolution soil moisture retrievals from the synergy of SMOS & MSG SEVIRI spaceborne observations. *Remote Sens. Environ.* **2016**, *180*, 403–471. [[CrossRef](#)]
5. Cambardella, C.; Moorman, T.; Parkin, T.; Karlen, D.; Novak, J.; Turco, R.; Konopka, A. Field-scale variability of soil properties in central Iowa soils. *Soil Sci. Soc. Am. J.* **1994**, *58*, 1501–1511. [[CrossRef](#)]

6. Western, A.W.; Grayson, R.B.; Blöschl, G. Scaling of soil moisture: A hydrologic perspective. *Annu. Rev. Earth Planet. Sci.* **2002**, *30*, 149–180. [[CrossRef](#)]
7. Petropoulos, G.P.; Srivastava, P.K.; Piles, M.; Pearson, S. EO-based operational estimation of soil moisture and evapotranspiration for agricultural crops in support of sustainable water management. *Sustainability* **2018**, *10*, 181. [[CrossRef](#)]
8. Kourgialas, N.N.; Koubouris, G.C.; Dokou, Z. Optimal irrigation planning for addressing current or future water scarcity in Mediterranean tree crops. *Sci. Total Environ.* **2019**, *654*, 616–632. [[CrossRef](#)]
9. Kourgialas, N.N.; Anyfanti, I.; Karatzas, G.P.; Dokou, Z. An integrated method for assessing drought prone areas—Water efficiency practices for a climate resilient Mediterranean agriculture. *Sci. Total Environ.* **2018**, *625*, 1290–1300. [[CrossRef](#)]
10. Snepvangers, J.; Heuvelink, G.; Huisman, J. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma* **2003**, *112*, 253–271. [[CrossRef](#)]
11. Kourgialas, N.N.; Karatzas, G.P. A flood risk decision making approach for Mediterranean tree crops using GIS; Climate change effects and flood-tolerant species. *Environ. Sci. Policy* **2016**, *63*, 132–142. [[CrossRef](#)]
12. Kourgialas, N.N.; Karatzas, G.P. A hydro-sedimentary modeling system for flash flood propagation and hazard estimation under different agricultural practices. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 625–634. [[CrossRef](#)]
13. Stein, A.; Hoogerwerf, M.; Bouma, J. Use of soil-map delineations to improve (co-) kriging of point data on moisture deficits. *Geoderma* **1988**, *43*, 163–177. [[CrossRef](#)]
14. Corradini, C. Soil moisture in the development of hydrological processes and its determination at different spatial scales. *J. Hydrol.* **2014**, *516*, 1–5. [[CrossRef](#)]
15. Petropoulos, G.P.; Griffiths, H.; Dorigo, W.; Xaver, A.; Gruber, A. Surface soil moisture estimation. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*; CRC Press: Boca Raton, FL, USA, 2013; pp. 29–48.
16. Srivastava, P.K. Satellite soil moisture: Review of theory and applications in water resources. *Water Resour. Manag.* **2017**, *31*, 3161–3176. [[CrossRef](#)]
17. Petropoulos, G.P.; McCalmont, J.P. An operational in situ soil moisture & soil temperature monitoring network for West Wales, UK: The WSMN network. *Sensors* **2017**, *17*, 1481.
18. Collins, F.; Bolstad, P. A comparison of spatial interpolation techniques in temperature estimation. In Proceedings of the NCGIA Third International Conference. Workshop on Integrating GIS and Environmental Modelling (CD-ROM), NCGIA, Santa Barbara, CA, USA, 21–26 January 1996.
19. Bárdossy, A.; Lehmann, W. Spatial distribution of soil moisture in a small catchment. Part 1: geostatistical analysis. *J. Hydrol.* **1998**, *206*, 1–15. [[CrossRef](#)]
20. Burrough, P.A.; McDonnell, R.A. Creating continuous surfaces from point data. In *Principles of Geographic Information Systems*; Oxford University Press: Oxford, UK, 1998.
21. Zhou, F.; Guo, H.-C.; Ho, Y.-S.; Wu, C.-Z. Scientometric analysis of geostatistics using multivariate methods. *Scientometrics* **2007**, *73*, 265–279. [[CrossRef](#)]
22. Yao, X.; Fu, B.; Lü, Y.; Sun, F.; Wang, S.; Liu, M. Comparison of four spatial interpolation methods for estimating soil moisture in a complex terrain catchment. *PLoS ONE* **2013**, *8*, 54660. [[CrossRef](#)] [[PubMed](#)]
23. Chai, X.; Shen, C.; Yuan, X.; Huang, Y. Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. *Geoderma* **2008**, *148*, 159–166. [[CrossRef](#)]
24. Ding, Y.; Wang, Y.; & Miao, Q. Research on the spatial interpolation methods of soil moisture based on GIS. In Proceedings of the International Conference on Information Science and Technology, Nanjing, China, 26–28 March 2011; pp. 709–711.
25. Fang, K.K.; Li, H.K.; Wang, Z.K.; Du, Y.F.; Wang, J. Comparative analysis on spatial variability of soil moisture under different land use types in orchard. *Sci. Hort.* **2016**, *207*, 65–72. [[CrossRef](#)]
26. Ford, T.W.; Quiring, S.M. Comparison and application of multiple methods for temporal interpolation of daily soil moisture. *Int. J. Climatol.* **2014**, *34*, 2604–2621. [[CrossRef](#)]
27. Chen, H.; Fan, L.; Wu, W.; Liu, H.B. Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought. *Environ. Monit. Assess.* **2017**, *189*, 525. [[CrossRef](#)]
28. Lu, G.Y.; Wong, D.W. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. [[CrossRef](#)]
29. Dobriyal, P.; Qureshi, A.; Badola, R.; Hussain, S.A. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydrol.* **2012**, *458*, 110–117. [[CrossRef](#)]

30. Verstraeten, W.W.; Veroustraete, F.; Feyen, J. Assessment of evapotranspiration and soil moisture content across different scales of observation. *Sensors* **2008**, *8*, 70–117. [[CrossRef](#)] [[PubMed](#)]
31. Patel, D.P.; Srivastava, P.K. Flood hazards mitigation analysis using remote sensing and GIS: Correspondence with town planning scheme. *Water Resour. Manag.* **2013**, *27*, 2353–2368. [[CrossRef](#)]
32. Otukey, J.; Blaschke, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S27–S31. [[CrossRef](#)]
33. Petropoulos, G.P.; Kontoes, C.; Keramitsoglou, I. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using support vector machines. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 70–80. [[CrossRef](#)]
34. Elatawneh, A.; Kalaitzidis, C.; Petropoulos, G.P.; Schneider, T. Evaluation of diverse classification approaches for land use/cover mapping in a Mediterranean region utilizing Hyperion data. *Int. J. Digit. Earth* **2014**, *7*, 194–216. [[CrossRef](#)]
35. Laslett, G.M. Kriging and splines: an empirical comparison of their predictive performance in some applications. *J. Am. Stat. Assoc.* **1994**, *89*, 391–400. [[CrossRef](#)]
36. Isaaks, E.H.; Srivastava, M.R. *An Introduction to Applied Geostatistics*; Oxford University Press: New York, NY, USA, 1989; p. 561.
37. Kourgialas, N.N.; Koubouris, G.C.; Karatzas, G.P.; Metzidakis, I. Assessing water erosion in Mediterranean tree crops using GIS techniques and field measurements: The effect of climate change. *Nat. Hazards* **2016**, *83*, 65–81. [[CrossRef](#)]
38. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*; Wiley: Melbourne, Australia, 2001.
39. Hutchinson, M. Thin plate spline interpolation of mean rainfall: Getting the temporal statistics correct. In *GIS and Environmental Modeling: Progress and Research Issues*; Wiley: Melbourne, Australia, 1996; pp. 85–90.
40. Lam, N.S.-N. Spatial interpolation methods: A review. *Am. Cartogr.* **1983**, *10*, 129–150. [[CrossRef](#)]
41. Dokou, Z.; Kourgialas, N.N.; Karatzas, G.P. Assessing groundwater quality in Greece based on spatial and temporal analysis. *Environ. Monit. Assess.* **2015**, *187*, 774. [[CrossRef](#)] [[PubMed](#)]
42. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press on Demand: Oxford, UK, 1997.
43. Kerry, R.; Oliver, M. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma* **2007**, *140*, 383–396. [[CrossRef](#)]
44. Marchant, B.P.; Lark, R.M. Optimized sample schemes for geostatistical surveys. *Math. Geol.* **2007**, *39*, 113–134. [[CrossRef](#)]
45. Armstrong, J.S.; Collopy, F. Error measures for generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast.* **1992**, *8*, 69–80. [[CrossRef](#)]
46. Bourennane, H.; King, D. Using multiple external drifts to estimate a soil variable. *Geoderma* **2003**, *114*, 1–18. [[CrossRef](#)]
47. Srivastava, P.K.; Islam, T.; Gupta, M.; Petropoulos, G.; Dai, Q. WRF dynamical downscaling and bias correction schemes for NCEP estimated hydro-meteorological variables. *Water Resour. Manag.* **2015**, *29*, 2267–2284. [[CrossRef](#)]
48. Srivastava, P.K.; Han, D.; Ramirez, M.R.; Islam, T. Machine learning techniques for downscaling smos satellite soil moisture using modis land surface temperature for hydrological application. *Water Resour. Manag.* **2013**, *27*, 3127–3144. [[CrossRef](#)]
49. Srivastava, P.K.; Han, D.; Rico-Ramirez, M.A.; Bray, M.; Islam, T. Selection of classification techniques for land use/land cover change investigation. *Adv. Space Res.* **2012**, *50*, 1250–1265. [[CrossRef](#)]
50. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–147. [[CrossRef](#)]

