



## Comparative study of interpolation methods for mapping soil pH in the apple orchards of Murree, Pakistan

Humair Ahmed<sup>1</sup>, Muhammad Tariq Siddique<sup>2</sup>, Muhammad Iqbal<sup>3</sup> and Fayyaz Hussain<sup>1</sup>

<sup>1</sup> Land Resources Research Institute, NARC, Islamabad

<sup>2</sup> PMAS-Arid Agriculture University Rawalpindi

<sup>3</sup> Space Applications and Research Complex, SUPARCO Islamabad

### Abstract

Soil pH is considered as a core indicator for nutrient bioavailability. Prevailing alkaline pH due to calcareousness in Pakistan is considered as one of the limiting factor for nutrient availability to plants. Exploring the spatial variability of soil variables serves as scientific basis for the generation of soil management strategies. Selection of best interpolation method to predict the soil properties at un-sampled locations is an important issue in the site specific investigations. This article evaluates Inverse distance weighting, global and local polynomial interpolation, radial basis function and kriging to determine the optimal interpolation method for mapping soil pH. Performance of the interpolation methods was analyzed using soil test (pH) data from 180 surface soil samples collected from 30 representative orchards grown in tehsil Murree. For inverse distance weighting, powers of 1, 2 and 3 were used and the number of neighbors for all methods ranged from 15 to 25. The conclusion of our study suggested that increased power of inverse distance weighting resulted in an increase in the prediction accuracy. Local polynomial interpolation method was more suitable as compared to global polynomial interpolation. Radial basis function with regularized and spline tension gave equivalent prediction accuracy. Higher errors (mean and mean absolute errors) were observed in case of ordinary kriging as compared to other interpolation methods. Digital maps generated by the higher powers of inverse distance weighting, local polynomial interpolation, and radial basis function were of higher accuracy.

**Keywords:** GIS, Geo-statistics, pH, Soil, Interpolation, Mapping

### Introduction

Digital Mapping has been widely used to develop a numerical or statistical model of various environmental variables and soil attributes for the spatial management of soils. Spatial management of soils is associated application system means assuming that application are made only in amounts and locations where they are needed i.e. targeted application of inputs. Economically variable rate system allows fertilizer money to be spent on the areas where response is expected and to be saved where response is unlikely (Bhatti, 2003). Among the physico-chemical properties of soil, significance of pH as nutrient bioavailability indicator cannot be underestimated. It has substantial influence on solubility and bioavailability of nutrients.

The micronutrients, B, Cu, Fe, Mn and Zn with the exception of Mo are more soluble at low pH (5.0-6.5). Concentrations of Fe<sup>+2</sup> and Mn<sup>+2</sup> may reach toxic levels at pH < 5.0. The Mo is more available at high pH i.e., 7-10 (Ahmed *et al.*, 2016). The bioavailability of Zn decreases hundred folds for each unit increase in soil pH in the range

4 to 9 (Lindsay, 1972; Ahmed *et al.*, 2010). Cations such as Ca<sup>+2</sup>, Mg<sup>+2</sup>, Na<sup>+</sup>, and K<sup>+</sup> exist in higher concentration in arid soil and are not soluble in neutral and alkaline pH range. The solubility of phosphorous (P) is optimum over a narrow pH range i.e., 6.5-7.5 (Lucas and Knezek, 1972). Continual spatial variability of physico-chemical properties and nutrients is one of the important characteristic features of the soil (Ahmed *et al.*, 2014). Observation of changes in physico-chemical properties from each and every corner of the area is impossible due to strenuous soil sampling and analyses (Jin *et al.*, 2012). Classical statistics helps to summarize and infer on the basis of information provided by the data set (Jin *et al.*, 2012).

Geo-statistics quantifies the spatial variability of nutrients; estimate the values of un-sampled locations (Ahmed *et al.*, 2014; Ahmed *et al.*, 2012). Various geo-statistical methods used for the prediction of soil variables from un-sampled location include inverse distance weighting, global and local polynomial interpolation, radial basis function and Kriging. The main assumption in inverse distance weighting method is that each input point has a local influence that lessens with increasing distance. Inverse

\*Email: humair\_uaar@yahoo.com

distance weighting produces maps by launching a neighborhood search of points and weighting these points by a power function. Increasing the power of inverse distance weighting often enhances the effect of points those are farther. Weights are more uniformly distributed between the neighboring points when lesser power is used for map generation (Poshtmasari *et al.*, 2012). The advantage of inverse distance weighting is that it is spontaneous and works well with evenly distributed points and it is sensitive to outliers. Unevenly distributed data clusters result in introduced errors (Balakrishnan *et al.*, 2011). The basic equation for interpolation by inverse distance weighting at an un-sampled location is given by Equation 1 as described by Gunarathna *et al.* (2016).

$$Z_j = \frac{\sum_i \left( \frac{Z_i}{d_{ij}^n} \right)}{\sum_i \frac{1}{d_{ij}^n}} \quad \text{Equation 1}$$

Where:  $Z_j$  indicates the estimated value for the unknown point at location  $j$ .  $d_{ij}$  indicates distance between known point  $i$  and unknown point  $j$ .  $Z_i$  manifests value at known point  $i$  while  $n$  indicates the user-defined exponent for weighting. Global polynomial interpolation fits a polynomial in the process of generating maps while local polynomial interpolation fits many polynomials, each within specified overlapping neighborhoods. The search neighborhood can be defined by using the size and shape, number of neighbors, and sector configuration (ESRI, 2011). Radial basis function methods are exact interpolation procedure and the generated maps from these methods go through each measured sample value. Each Radial basis function has a different shape and results in a slightly different interpolation map (Poshtmasari *et al.*, 2012). Radial basis functions are typically used to build up function approximations of the form given in Equation 2 (ESRI, 2011).

$$y(x) = \sum_{i=1}^n w_i \phi(\|x - x_i\|) \quad \text{Equation 2}$$

Where approximating function  $y(x)$  is represented as a sum of  $N$  radial basis functions, each associated with different centre  $x_i$  weighted by an appropriate coefficient  $w_i$ . The weights  $w_i$  is estimated using matrix methods of linear least square. Splines consist of polynomials, which describe pieces of a line or surface, and they are fitted together so that they join smoothly (Webster & Oliver, 2001). The essence of kriging is the weighted average that takes in to account the known spatial dependence expressed by the semivariogram. It appears to provide more accurate description of spatial structure for studied soil properties than other interpolations as demonstrated in many comparative studies (Kravchnko, 1999). The basic equation

for interpolation by kriging at an unsampled location is given by Equation 3 (ESRI, 2011).

$$Z_k(X_0) = \sum_{i=1}^{n(h)} \lambda_i Z(X_i) \quad \text{Equation 3}$$

Where  $Z_k(X_0)$  = Weighted average of the observed value,  $\lambda_i$  = weights associated with the data points, which takes in account geographic relationship of the sampled points. Points near ( $X_0$ ) carry more weight than distant points; points close together tend to have a single while lone points carry full weight. Several assumptions are made in kriging. One of the assumptions is stationarity of the data. There are two kinds of stationarity: First order stationarity, which assumes that the mean remains constant regardless the location. Second order stationarity means that the covariance exists and it depends only on the lag distance  $h$  but not on the sampling position (Nielsen, 2003).

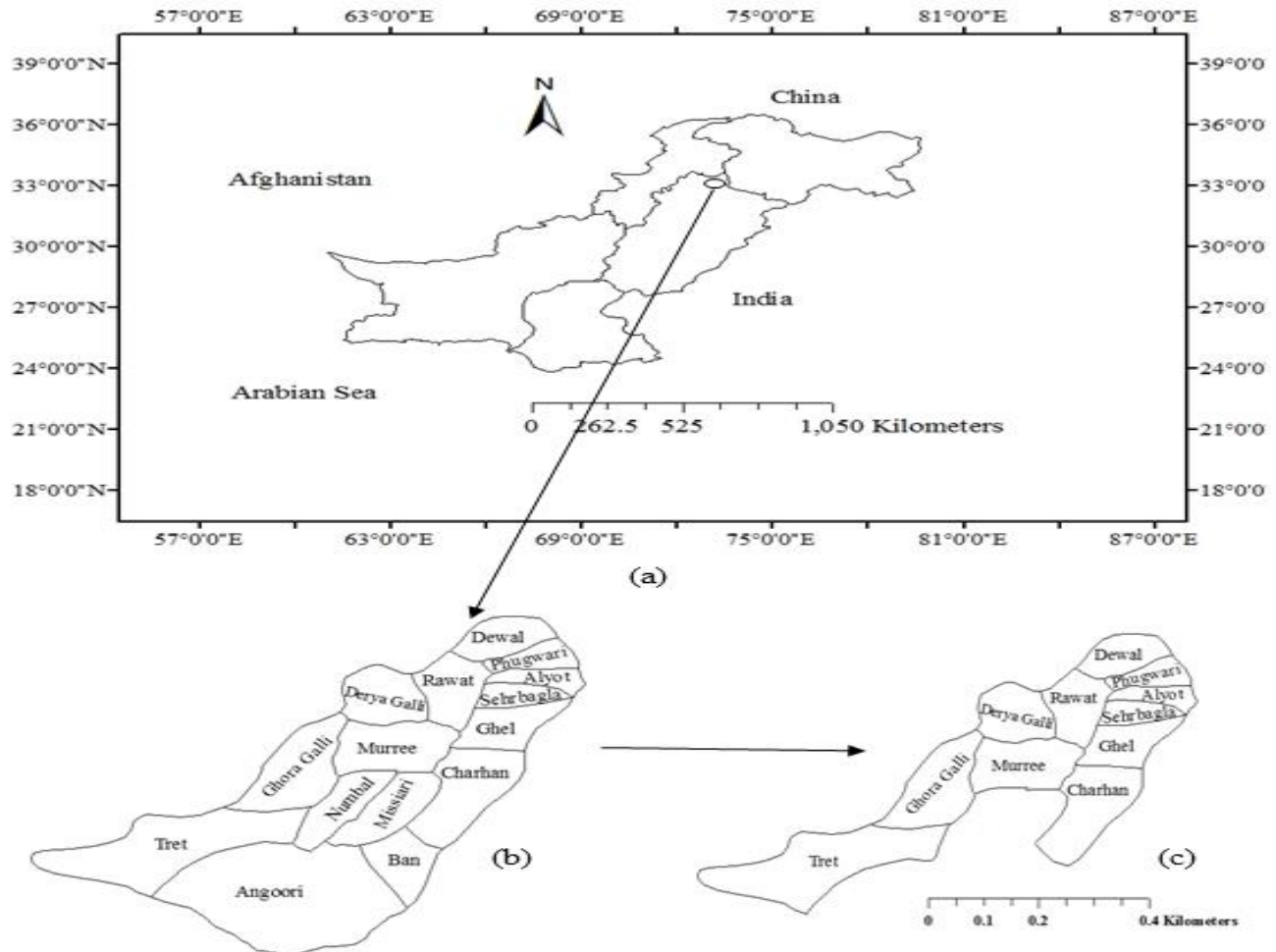
There is no single preferred method for data interpolation. In order to determine the spatial variability of soil properties the best interpolation method needs to be identified which is stimulated in various comparative studies in various regions (Zimmerman *et al.*, 1999). The main objective of this study was to review and evaluate common interpolation methods, i.e., inverse distance weighting, global and local polynomial interpolation methods, radial basis function and ordinary kriging and to evaluate the accuracy of generated maps of soil pH property with suitable methods.

## Material and Methods

The study area was conducted at Murree area is located between the  $33^{\circ} 54' 26.56''$  N and  $73^{\circ} 23' 37.68''$  E on the north east of Islamabad (capital city of Pakistan) and spread over the  $434 \text{ km}^2$  (Figure 1). Climatic conditions are congenial for growing apple orchards. Six georeferenced soil samples (0-15 cm) were collected at  $6 \text{ m}^2$  from 30 apple orchards to cover the whole apple producing area of Murree. Soil samples were air dried, crushed, and passed through a 2 mm sieve and were stored in plastic bags for further analysis. The samples were analyzed for pH (McLean, 1982). Co-efficient of variance was used to examine the variability of pH within the dataset. Variables having CV% values  $< 15\%$  are grouped as least variable whereas those having CV between 15 to 35% are categorized as moderately variable. Co-efficient of variance value more than 35% indicates high variability (Wilding, 1985).

Collected spatial data were used for geo-referencing and digitization of the base maps gained from Survey of





**Figure 1: Geographical location of the study area: (a) Pakistan (b) Murree map indicating all Union Councils (c) Map indicating survey area**

Pakistan. Attribute data were used for geostatistical analyses. Generation of digital maps was accomplished by simulation of geo-statistically analyzed data within the boundaries of digitized maps. The validation and the fitness of the interpolation method were tested via a method called cross validation. Cross validation estimation is obtained by leaving one sample out and using the remaining data to estimate the value. This test allows evaluating the goodness of fit of the method and the appropriateness of neighborhood, while the interpolation values are compared to the real (Leuangthong *et al.*, 2004). The efficacy of applied techniques was evaluated using mean error calculated by the formula (Equation 4) and mean absolute error (MAE) calculated by the formula (Equation 5) and RMSE calculated by the formula (Equation 6).

$$ME = \frac{1}{n} \sum_{i=0}^n [Z^*x_i - Zx_i] \quad \text{Equation 4}$$

$$MAE = \frac{1}{n} \sum_{i=0}^n [ |Z^*(x_i) - Z(x_i)| ] \quad \text{Equation 5}$$

$$RMSE = [ \sum_{i=0}^n (Z^*x_i - Zx_i)^2 / N ]^{\frac{1}{2}} \quad \text{Equation 6}$$

Where  $Z^*(x_i)$  and  $Z(x_i)$  represent the observed and predicted values respectively.

## Results and Discussion

Descriptive statistics of the data set related to soil pH is summarized in the Table 1. Soil pH ranged from 6.35 to 8.25 with the mean value of  $7.51 \pm 0.49$ . Skewness (-0.075) kurtosis (-0.83), Frequency distribution (Figure 2) and normal QQ plot (Figure 3) indicated the normal distribution of the data. Soil pH was slightly alkaline to strongly alkaline in nature having least heterogeneity as the CV value was nearly equal to 15%. Soil pH indicated that soils



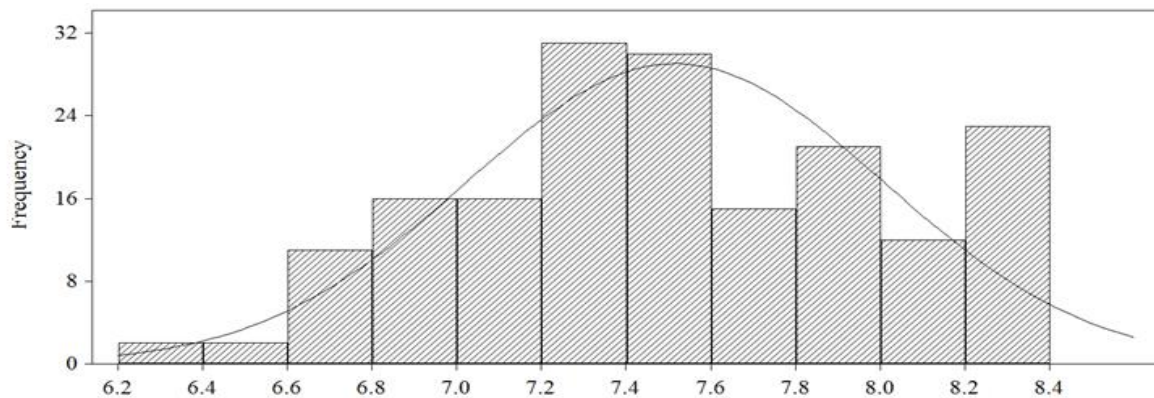
of the sampled orchards were free from the salinity or sodicity hazard.

**Table 1: Summary statistics of soil pH in the surveyed apple orchards**

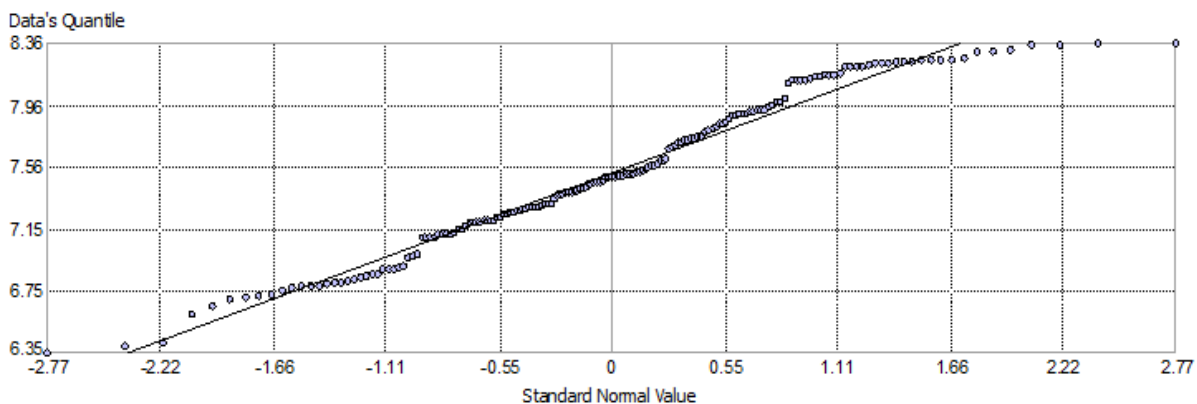
Statistical Parameter	Values	Statistical Parameter	Value
Mean	7.51	Minimum	6.35
Standard Deviation	0.49	Maximum	8.36
*CV (%)	15.32	Skewness	-0.075

\*Coefficient of variance

a decrease of mean, root mean square and mean absolute error values with application of inverse distance weighting smooth method. Hence our results indicated that increasing power of inverse distance weighting, standard and smooth technique enhanced the accuracy in the prediction and map generation. Increase in accuracy for map generation due enhanced power was due to the main assumption in inverse distance weighting method which states that each input point has a local influence that lessens with increasing distance because inverse distance weighting produces maps by launching a neighborhood search of points and weighting these points by a power function. Increasing



**Figure 2: Frequency distribution of soil pH dataset of surveyed apple orchards**



**Figure 3: Normal QQ plot indicating normal distribution of the data of surveyed orchards**

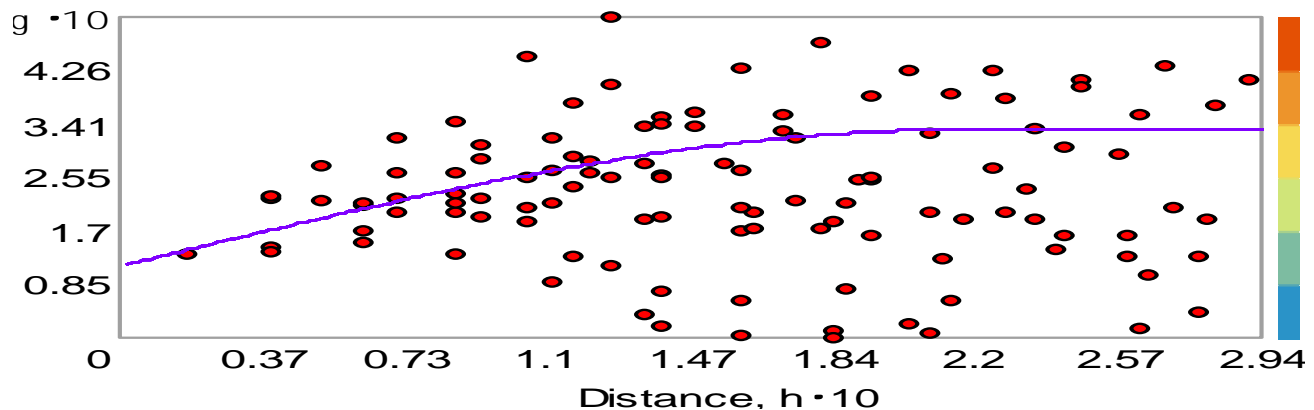
Our results indicated that when inverse distance weighting was applied to the data set using standard and smooth type having various powers a slight difference in case of root mean square error was observed. Increasing power resulted in a gradual decrease in the mean, root mean square and mean absolute error respectively when inverse distance weighting standard technique was evaluated (Table 1). Similarly increasing power (1, 2 and 3) also resulted in

power of inverse distance weighting often enhances the effect of the points that are farther from one another. Weights are more uniformly distributed between the neighboring points when lesser power is used for map generation (Poshtmasari *et al.*, 2012). Global polynomial interpolation indicated higher mean error, root mean square error and mean absolute error as compared to local polynomial interpolation method (Table 2).

**Table 2: Parameters related to interpolation techniques applied to soil pH in the surveyed orchards**

Prediction Model Details		Prediction Errors		
Method	Type	Mean Error	MAE <sup>s</sup>	RMSE <sup>Ω</sup>
	<b>Standard (Power 1)</b>	<b>- 0.0160</b>	<b>0.2269</b>	<b>0.29</b>
IDW <sup>*</sup>	Standard (Power 2)	0.0050	0.1952	0.25
	Standard (Power 3)	0.0020	0.1769	0.23
	Smooth (Power 1)	0.0230	0.2319	0.38
	Smooth (Power 2)	0.0008	0.1952	0.29
	Smooth (Power 3)	0.0017	0.1873	0.24
PI <sup>+</sup>	Global	0.0010	0.3820	0.44
	Local	0.0183	0.1684	0.22
RBF <sup>ε</sup>	Regularized	0.00314	0.1278	0.18
	Tension	0.00330	0.1378	0.19
Ordinary Kriging	Spherical	0.00713	0.2211	0.28

<sup>s</sup> Mean Absolute Error <sup>Ω</sup> Root Mean Square Error <sup>\*</sup> Inverse Distance Weighting <sup>+</sup> Polynomial Interpolation, <sup>ε</sup> Radial Basis Function

**Figure 4: Semivariogram model for soil pH**

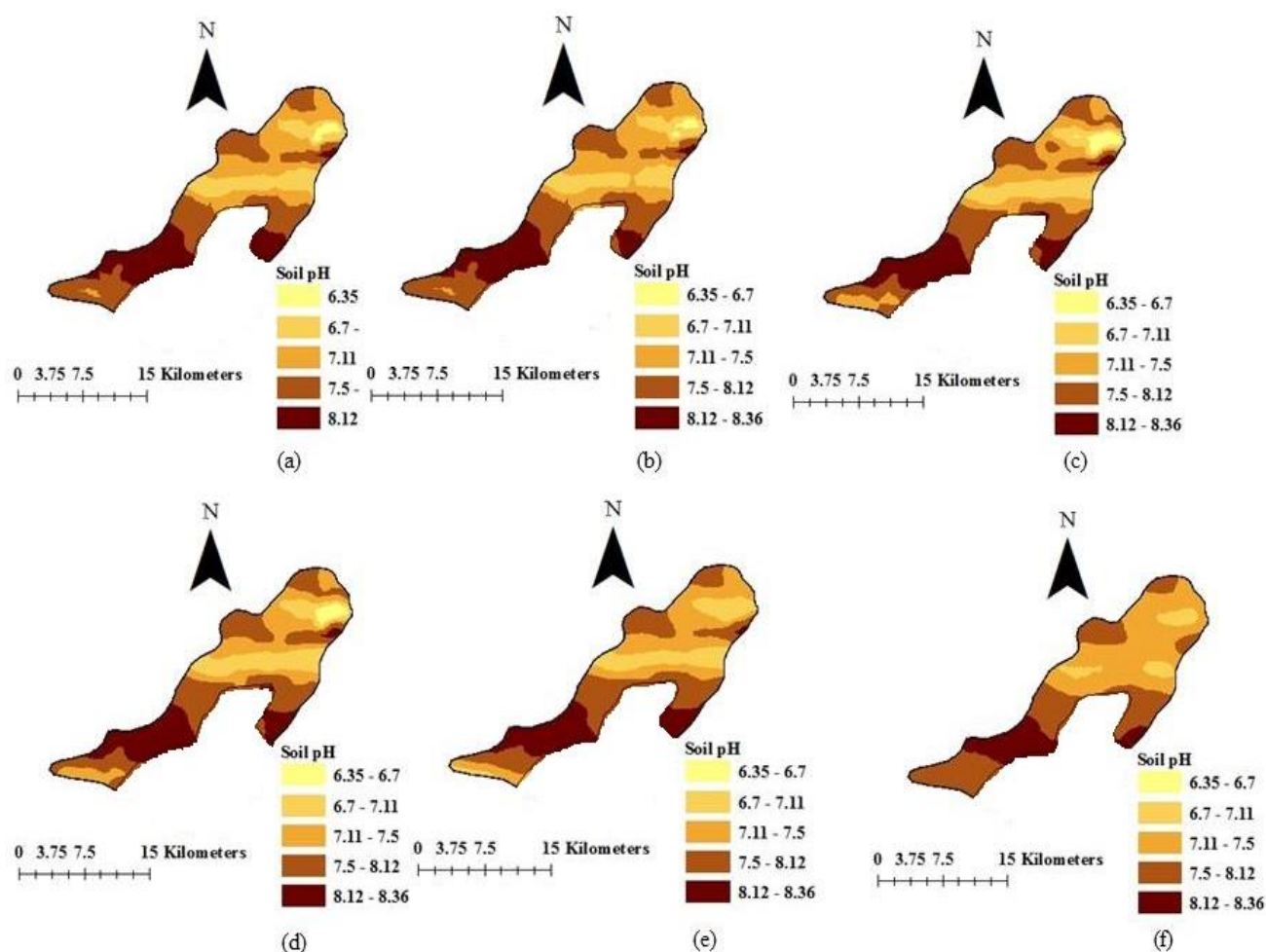
Global polynomial interpolation fits single polynomial while local polynomial interpolation fits many polynomials, each within specified overlapping neighborhoods (ESRI, 2011). Therefore, the use of many polynomials with in overlapping neighborhood might be the reason for lower mean error, root mean square error and mean absolute error values in the local polynomial interpolation method. Radial basis function using regularized and tension Splines gave similar results in the prediction of soil pH as root mean square error observed in the case of regularized and tension spline was 0.18 and 0.19 respectively (Table 2). When ordinary kriging was applied to interpolate the data using spherical model, Semivariogram (Figure 4) indicated lower nugget to sill ratio, Moreover rising of semivariogram to a certain point and its levelling off later on also indicated a strong spatial dependence of data related to soil pH. Averaged control points were used to examine the spatial autocorrelation as less cluttered view of the spatial autocorrelation in the data and showed smoother changes in

the semivariogram values than the binned points. Our results indicated slightly higher mean, mean absolute and root mean square error in case of kriging as compared to other evaluated interpolation methods which manifested that ordinary kriging was less suitable method for interpolation of soil pH in the surveyed area. However kriged map was generated as this technique is most commonly utilized by the soil scientists (Mulla *et al.*, 1992; Bhatti *et al.*, 1998a; Takata *et al.*, 2007; Takata *et al.*, 2008; Ahmed *et al.*, 2012, Ahmed *et al.*, 2014). Generated maps (Figure 5) using through methods which indicated equivalent accuracy according to cross validation indicated similar maps while ordinary kriging without any data transformation indicated different maps as compared to other methods.

## Conclusion

All selected interpolation methods indicated a good potential for interpolation of soil pH. However inverse





**Figure 5: Spatial distribution of soil pH interpolated by using various techniques**

a) Inverse distance weighting (Standard), b) Inverse distance weighting (smooth), c) Local polynomial interpolation, d) Radial basis function regularized spline, e) Radial basis function tension spline, f) Ordinary kriging

distance weighting standard and smooth using power 3 offered lowest mean error, mean absolute error and root mean square error. Local polynomial method for interpolation of soil pH was observed efficient as compared to global polynomial interpolation. Radial basis function methods using regularized and tension spline exhibited equivalent results for prediction of soil pH. Ordinary kriging with the spherical model proved to be a less suitable method for predicting soil pH. Hence contour map generation using inverse distance weighting, standard with power three, local polynomial interpolation, and radial basis function method using regularized and tension spline was used to prepare promising contour maps for the prediction of soil pH from un-sampled locations. Although mean, mean absolute and root mean square error values obtained from ordinary kriging were higher but contour maps were generated to observe the difference. Maps generated by

inverse distance weighting, local polynomial interpolation and radial basis function with regularized and tension spline manifested similar patterns while maps generated through ordinary kriging showed deviation in the patterns demonstrated by all other tested methods. Evaluation of interpolation methods before making prediction about soil attributes through GIS/geo-statistics is highly recommended.

## References

- Ahmad, H., M. T. Siddique, I. A. Hafiz and Ehsan-ul-Haq. 2010. Zinc status of apple orchards and its relationship with selected physico-chemical properties in Murree tehsil. *Soil and Environment* 29: 142-147.
- Ahmed, H., M. T. Siddique, S. Ali, N. A. Abbasi, A. Khalid and R. Khalid. 2014. Micronutrient indexing in the apple orchards of Northern Punjab, Pakistan using

- geostatistics and GIS as diagnostic tools. *Soil and Environment* 33: 07–16.
- Ahmed, H., M. T. Siddique, S. Ali, A. Khalid and N. A. Abbasi. 2012. Mapping of Fe and impact of selected physico-chemical properties on its bioavailability in the apple orchards of Murree region. *Soil and Environment* 31: 100–107.
- Ahmed, H., M. T. Siddique, F. Hussain, A. Hayat, M. J. Atif and N. Saleem. 2016. GIS based mapping of copper, manganese and boron availability in apple orchards of Murree, Pakistan. *Science, Technology and Development* 35 : 160-170.
- Balakrishnan, P., A. Saleem, and N. D. Mallikarjun. 2011. Groundwater quality mapping using geographic information system (GIS): A case study of Gulbarga City, Karnataka, India. *African Journal of Environmental Science and Technology* 5: 1069-1084.
- Bhatti, A.U. 2003. Spatial variability and its management in agriculture. Higher Education Commission of Pakistan. p. 86-91.
- Bhatti, A.U., F. Hussain, Farmanullah and M.J. Khan. 1998a. Use of spatial patterns of soil properties and wheat yields in geostatistics for determination of fertilizer rates. *Communications in Soil Science and Plant Analysis* 29: 509-522.
- ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- Gunarathna, M.H.J.P., M.K.N. Kumari and K.G.S. Nirmanee. 2016. Evaluation of interpolation methods for mapping pH of groundwater. *International Journal of Latest Technology in Engineering, Management & Applied Science* 3: 1-5.
- Jin, J.W., H.C. Ye, Y.F. Xu, C.Y. Shen and Y.F. Huang. 2012. Spatial and temporal pattern of soil fertility, quality and analysis of related factors in urban-rural transition zone of Beijing. *Advances in Environmental Biology* 6: 1620–1625.
- Kravchenko, A., and D. G. Bullock. 1999. A Comparative study of interpolation methods for mapping soil properties. *Agronomy Journal* 91:393-400.
- Lucas, E. and Knezek. Climate and soil conditions promoting micronutrient deficiencies in plants. 1972. p. 265-287. In: *Micronutrients in agriculture*. Mortvedt, J. J., P. M. Giordano and W. L. Lindsay, (eds). Soil Science Society of America, Inc. Madison, Wisconsin USA.
- Leuangthong O., J.A. McLennan and C.V. Deutsch. 2004. Minimum acceptance criteria for geostatistical realizations. *Natural Resources Research* 13:131–141.
- Lindsay, W.L. Inorganic phase equilibria of micronutrients in soils. 1972. p. 41-57. In: *Micronutrients in Agriculture*. J.J. Mortvedt, P.M. Giordano and W.L. Lindsay (eds.). Soil Science Society of America, Inc. Madison, Wisconsin USA.
- McClean, E.O. 1982. Soil pH and lime requirement. p. 198–209. In: *Methods of Soil Analysis, Part 2: Chemical and Microbiological Properties*. A.L. Page, R.H. Miller and D.R. Keeney (eds.). American Society of Agronomy. Madison, WI, USA.
- Mulla, D.J., A.U. Bhatti, M.W. Hammond and J.A. Benson. 1992. A comparison of winter wheat yield and quality under uniform versus spatially variable fertilizer management. *Agriculture, Ecosystem and Environment* 38: 301-311.
- Nielsen, D. R. and O. Wendroth. 2003. Spatial and temporal statistics – sampling field soils and their vegetation. *Journal of Plant Nutrition and Soil Science* 166 : 667-668.
- Poshtmasari, H.K., Sarvestani, Z.T., Kamkar, B., Shataei, S. and Sadeghi, S. 2012. Comparison of interpolation methods for estimating pH and EC in agricultural fields of Golestan province (north of Iran). *International Journal of Agricultural and Crop Sciences* 157-167.
- Takata, Y., S. Funakawa, K. Akshalov, N. Ishida, and T. Kosaki. 2007. Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information. *Soil Science and Plant Nutrition* 53: 289-299.
- Takata, Y., S. Funakawa, J. Yanai, A. Mishima, K. Akshalov, N. Ishida, and T. Kosaki. 2008. Influence of crop rotation system on the spatial and temporal variation of the soil organic carbon budget in northern Kazakhstan. *Soil Science and Plant Nutrition* 54: 159-171.
- Webster, R and M. Oliver. 2001. *Geostatistics for Environmental Scientists*. John Wiley and Sons.
- Wilding, L.P. Spatial variability: Its documentation, accommodation and implication to soil surveys. 1985. p. 166-194. In: *Soil Spatial Variability*. D.R. Nielsen and J. Bouma (eds.). Pudoc, Wageningen, Netherland.
- Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology* 31: 375-390.

