Identify effective auxiliary variables in cokriging method to estimate spatial variability of infiltration rate

A. Yekzaban\textsuperscript{a,*}, M. Shabanpour\textsuperscript{a}, N. Davatgar\textsuperscript{b}
\textsuperscript{a}Department of Soil Science, Faculty of Agriculture, University of Guilan, Rasht, IRAN.
\textsuperscript{b}Department of Soil and Water Research, Rice Research Institute of Iran (IRRI), Rasht, IRAN.

*Corresponding author; Department of Soil Science, Faculty of Agriculture, University of Guilan, Rasht, IRAN.

**ARTICLE INFO**

**ABSTRACT**

Measurements of infiltration rate (IR) in the field are costly, time consuming, and relatively cumbersome. IR are sensitive to some soil properties, thus the cokriging with auxiliary variables can sometimes improve estimates for less density sampled primary variable. The objectives of this study were to determine the spatial relationships between IR and some soil properties affecting IR and to identify possibility of using cokriging method. Infiltration rate test were conducted using double ring infiltrometers until steady state. 75 field measured IR were obtained at a nearly regular grid spacing of 10 m. The correlation coefficient between IR and OM and silt were comparatively good. Semivariogram and cross-semivariogram of these variables with moderate to strong spatial dependence were fitted into the spherical model, range spatial dependence above mentioned soil properties were generally greater than 24 m. The cross validation analysis showed that both kriging and cokriging provided reasonable estimates for IR. Differences among kriging and cokriging with using OM as auxiliary variable were relatively small. However using silt content as auxiliary data for the estimation of IR in cokriging method was consistently more effective than kriging on IR alone, and could reduce prediction error by 15% as compared kriging method.

© 2014 Sjournals. All rights reserved.
1. Introduction

Infiltration is a dominant process controlling crop yields, solute transport, run off and soil erosion. Infiltration rate may vary from very low to very high because of variability in the soil physical characteristics. Soil infiltration rate is mostly affected by soil texture and soil moisture (Radcliffe and Rasmussen, 2000; Dingman, 2002). Moreover soil properties such as infiltration rate change in time and space continuously. There are several causes for these changes of soil properties such as pedologic soil formation factor, topography, vegetation, cultivation history and variability arising from uneven field management (Tesfahunegn et al., 2011).

Therefore developing suitable schemes for the design and management of irrigation and drainage systems requires sufficient and reliable data of soil and water properties such as infiltration rate (Alemi et al., 1988). However since measuring infiltration rate in a field is costly, time consuming, and cumbersome it is difficult to estimate infiltration rate values at unobserved site with an acceptable level of accuracy, especially when spatial variability of this property is high, thus this requires quantifying the spatial information for infiltration rate (Ersahin, 2003).

Using geostatistics is feasible to characterize and quantify the spatial variability of soil samples, perform rational interpolation and estimate variance between the point values sampled in the spatial field by regional theory (Zhang et al., 2011). Cokriging uses the spatial information on infiltration rate along with spatial correlation between infiltration rate and an auxiliary variable to make estimations on unobserved sites (Ersahin, 2003). In contrast to the limited availability of information on infiltration rate, data related to major soil characteristics are often much more readily available.

Many researchers had applied cokriging methods in determination of spatial variability of soil physical properties (Triantafilis et al., 2001; Meul and Van Mirvence, 2003; Millan et al., 2012) and soil chemical properties (Wu et al., 2003; Robinson and Metternicht, 2006). They concluded that cokriging was superior to kriging in minimizing estimation variance. For example Ersahin (2003) used kriging and cokriging methods for investigation of spatial variability of infiltration rate in turkey. The soil bulk density was used as an auxiliary variable in cokriging method in this study. Result illustrated that the cokriging method was a suitable technique for estimation of infiltration rate. Wu et al. (2006) used organic matter and pH as auxiliary variables to estimate DTPA-extractable soil Zn in northern North Dakota. They found that cokriging on Zn(DTPA) using OC and pH as auxiliary variables, was consistently more effective than kriging on Zn(DTPA) alone. Moreover cokriging with OC and pH together provided additional benefit. The objective of this study were (i) to quantify the spatial structure of infiltration rate and soil properties affecting this, (ii) to evaluate the auxiliary soil characteristics can be used to improve predictions of infiltration rate, when data for infiltration rate are not available.

2. Materials and methods

2.1. Description of study area

This study was conducted in a near flat agricultural field (2-3% Slope) of about 1ha (85 m×85 m) located in Rasht, north of Iran. The climate is temperate with mean annual precipitation of 1200mm, mostly falling in the autumn and spring. Minimum and maximum monthly mean temperatures were 6.6 and 25°C respectively. Conventional tillage was performed two years ago with a moldboard plow at a depth of 30 cm.

2.2. Soil sampling and analysis techniques

The field was intensively sampled on a nearly regular grid spacing of 10 m in October 2012 (Fig. 1). Soil samples were obtained from three points near each site within the 0-30 cm soil surface layer (plow pan). All 75 samples were analyzed in the laboratory for sand, silt, and clay contents by using the hydrometer method (Gee and Bauder, 1986), organic matter content (OM) was analyzed using the Walkley-Black method (Nelson and Sommers, 1982), mean weight diameter (MWD) of soil aggregates by using wet sieving (Kemper and Rosenau, 1986), infiltration rate (IR) tests were done with double-ring infiltrometers until final (steady state) IRs was reached (Klute and Dirksen, 1986) for each site before soil sampling. In addition soil porosity was determined from Particle density and Bulk density with pycnometer and cylinder methods, respectively (Jacob and Clark, 2002).
2.3. Statistical analyses

Data analyses for each variable were done in four steps: (i) normality tests were applied (Shapiro-Wilks) and in the variables, do not have normality distribution, log-transformation was performed; (ii) distributions were describe with classical statistics; (iii) correlation between infiltration rate and other soil properties were determined; (iv) soil properties highly (significant) correlated with infiltration rate were selected as potential auxiliary variable for use in the cokriging procedure.

2.4. Geostatistical analysis

Before cokriging, the spatial variability of correlated variables was modeled with the aid of semivariograms and cross-semivariograms, which was determined to ascertain the degree of spatial variability. A semivariogram shows auto-correlation as a function of distance, which was defined as following equation,

$$\gamma(h) = \frac{1}{2N(h)} \sum_{k=1}^{N(h)} [(Z_i(x_k) - Z_j(x_k + h))(Z_j(x_k) - Z_j(x_k + h))]$$

Where, $\gamma$ is the semivariance (when $i = j$) with respect to random variable $Z_i$; $h$ is the separation distance; $N(h)$ is the number of pairs of $Z_i(x_k)$ and $Z_j(x_k)$ in a given logged distance interval of $(h + dh)$. When $i \neq j$, $\gamma$ is the cross-semivariance which is a function of $h$; and $Z_i(x_k)$ and $Z_j(x_k)$ are the observed principle and auxiliary values at $x_k$ location, respectively (Ersahin, 2003).

Several semivariogram and cross-semivariogram were evaluated to choose the best fit with the data. Spherical model were fitted to both experimental semivariograms and cross-semivariograms was done on the basis of regression ($r^2$) and residual sum square (RSS) (Robinson and Metternicht, 2006). The parameter of the model fitted to the experimental semivariograms and cross-semivariograms can be used with the data for prediction at points or over blocks. Cokriged predictions are a weighted average of the principle and auxiliary data, at the unknown point or block, so cokriging with secondary variable has the potential to improve estimates of primary variable, which was defined as following equation,

$$z^*(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z_i(x_k) + \sum_{j=1}^{m} \lambda_j \cdot z_j(x_k)$$

(2)
Where $Z^*(x_0)$ is the estimated value at $x_0$ location; $\lambda_i$ and $\lambda_j$ refers to weighing factors of principle and auxiliary variables, respectively; and $n$ and $m$ are the numbers of principle and auxiliary values (Isaaks and Srivastava, 1989).

Accuracy of interpolation was evaluated through a Cross-validation process using the Mean Absolute Error (MAE) (Millan et al., 2012) and General Standard Deviation (GSD) (Isaaks and Srivastava, 1989).

$$MAE = \frac{1}{n} \sum_{k=1}^{n} \left| z^* (x_k) - z(x_k) \right|$$

(3)

$$GSD = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left( z^* (x_k) - z(x_k) \right)^2 / z(x)}$$

(4)

Where $Z(x_k)$ and $Z^*(x_k)$ are the observed and estimated value at $x_k$ location, respectively; $\bar{Z}(x)$ is the mean of total observed values. The GS+ software package (Gamma Design Software, 2001) was used for performing geostatistical analysis.

3. Results and discussion

The descriptive statistics of the soil properties in the study field showed high skewness for some of the parameters (Table 1). The data for MWD, porosity and silt had low to moderate skewness, but data for IR, OM, clay and sand were far from normally distributed. For IR the SD is high and data were strongly positively skewed. Highly skewed parameters indicate that these variables have a local distribution. The reason for soil parameters being distributed abnormally may be associated with differences in management practices, land use, vegetation cover, and topographic effects (Tesfahunegn et al., 2011). Iqbal et al. (2005) reported that organic matter, clay and sand content were skewed significantly. Also log-normal distribution of the infiltration rate was reported by Sisson and Wieranga (1981) and Haws et al. (2004). A common log-transformation was successful in normalizing the data to reduce of skewness. By using the logarithm transformed variables we make ensure that the variables approximately have Gaussian distribution.

Variability in distribution of parameters is measured by coefficient of variation (CV). Base on the CV values, infiltration rate was the most variable soil measured parameter, with CV greater than 35%. Ersahin (2003) and Shukla et al. (2004) indicated that hydraulic properties showed the largest variability. The results are possibly related to the soil tillage practices caused the maximum change IR values at field measurement. MWD, OM, clay and sand were moderately variable, with CV between 15 and 35%, while porosity and silt were least variable (CV<15%).

<table>
<thead>
<tr>
<th>Soil Variable</th>
<th>Unit</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>CV</th>
<th>Correlation coefficients with IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
<td>cm h$^{-1}$</td>
<td>0.33</td>
<td>26.28</td>
<td>7.9</td>
<td>7.06</td>
<td>1.07*</td>
<td>0.07</td>
<td>88.49</td>
<td>1.000</td>
</tr>
<tr>
<td>Clay</td>
<td>%</td>
<td>22.00</td>
<td>48.00</td>
<td>32.0</td>
<td>5.29</td>
<td>0.61*</td>
<td>0.47</td>
<td>16.91</td>
<td>-0.153</td>
</tr>
<tr>
<td>Silt</td>
<td>%</td>
<td>51.00</td>
<td>70.00</td>
<td>59.7</td>
<td>4.33</td>
<td>0.34</td>
<td>-0.39</td>
<td>7.26</td>
<td>0.273</td>
</tr>
<tr>
<td>Sand</td>
<td>%</td>
<td>5.00</td>
<td>18.00</td>
<td>9.2</td>
<td>2.86</td>
<td>0.74*</td>
<td>0.15</td>
<td>31.10</td>
<td>0.054</td>
</tr>
<tr>
<td>Porosity</td>
<td>%</td>
<td>43.20</td>
<td>60.00</td>
<td>50.3</td>
<td>3.77</td>
<td>0.36</td>
<td>-0.20</td>
<td>7.50</td>
<td>0.174</td>
</tr>
<tr>
<td>MWD</td>
<td>mm</td>
<td>0.61</td>
<td>2.25</td>
<td>1.3</td>
<td>0.47</td>
<td>0.47</td>
<td>-0.06</td>
<td>0.83</td>
<td>31.20</td>
</tr>
<tr>
<td>OM</td>
<td>%</td>
<td>1.30</td>
<td>4.29</td>
<td>2.4</td>
<td>0.66</td>
<td>0.69*</td>
<td>0.28</td>
<td>26.86</td>
<td>0.305*</td>
</tr>
</tbody>
</table>

*Significant at the 0.05 probability level; Min: Minimum; Max: Maximum; SD: Standard deviation; CV: Coefficient of variation; IR: Infiltration rate; MWD: Mean weight diameter; OM: Organic matter.

The correlation among IR and other properties and the related level of significant are listed in Table 1. The correlation coefficients between soil properties investigated and IR were not statistically significant in all case. The weak correlation between IR and soil properties might be a hint to management process and root channels. Virtually, noting significant correlation between IR and other variables, indicating that these variables explained
different portions of the variability in IR. The linear correlation between IR and OM was 0.305 and significant (P<0.05), while the correlation between IR and silt was comparatively good. No significant correlation between IR and percentage of sand and OM and significant correlation between IR and silt were reported by Ersahin (2003). Reynolds and Zebchuk (1996) concluded that hydraulic properties were primarily affected by a well-developed and stable soil structure, and not by the soil texture, organic carbon, or surface topography. Rasse et al. (2000) found that differences in IR could be due to the presence of root channels and macro porosities. In other research Jorgensen et al. (2002) reported vertical continuity of the macro pore sequences would intuitively seem to be an important factor affecting IR.

Table 2

Semivariogram and cross-semivariogram parameters with prediction error values.

<table>
<thead>
<tr>
<th>Variable</th>
<th>model</th>
<th>C0</th>
<th>C0+C</th>
<th>Range (m)</th>
<th>R2</th>
<th>RSS</th>
<th>Class</th>
<th>C0/(C0+C)</th>
<th>MAE</th>
<th>GSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR (cm h⁻¹)</td>
<td>Spherical</td>
<td>0.2500</td>
<td>0.7700</td>
<td>24.0</td>
<td>0.95</td>
<td>2.5×10⁻³</td>
<td>M</td>
<td>32.4</td>
<td>4.87</td>
<td>0.77</td>
</tr>
<tr>
<td>Silt (%)</td>
<td>Spherical</td>
<td>3.3600</td>
<td>20.450</td>
<td>58.0</td>
<td>0.99</td>
<td>9.3×10⁻¹</td>
<td>S</td>
<td>16.4</td>
<td>2.42</td>
<td>0.05</td>
</tr>
<tr>
<td>OM (%)</td>
<td>Spherical</td>
<td>0.0313</td>
<td>0.0755</td>
<td>52.0</td>
<td>0.96</td>
<td>4.9×10⁻⁵</td>
<td>M</td>
<td>41.4</td>
<td>0.37</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Cross-Semivariogram</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IR / OM</td>
<td>Spherical</td>
<td>0.0060</td>
<td>0.0290</td>
<td>36.0</td>
<td>0.98</td>
<td>3.1×10⁻³</td>
<td>M</td>
<td>20.7</td>
<td>4.85</td>
<td>0.74</td>
</tr>
<tr>
<td>IR / Silt</td>
<td>Spherical</td>
<td>0.0018</td>
<td>0.0119</td>
<td>42.0</td>
<td>0.97</td>
<td>9.2×10⁻⁷</td>
<td>S</td>
<td>15.2</td>
<td>4.18</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 2 listed the semivariogram and cross-semivariogram parameters for the IR and OM and silt percentage that were the most related to IR. The cross-semivariogram between IR and other soil properties shows little spatial co-variability (data not shown), which echoes the lack of relationship and near zero correlation coefficient (Table 1). According to the higher regression (r²) and lower Residual sum square (RSS), experimental semivariogram and cross-semivariogram of soil properties were best fitted to spherical model (Fig. 2). Semivariance increased with distance between samples (lag distance) to a constant value (sill) at a given separation distance (the range of spatial dependence) for spherical model (Isaaks and Srivastava, 1989). To compare the spatial dependence of different variable can use the ratio of the nugget and the sill after fitting models. This ratio was used to define three classes of spatial dependence for measured variables (kilic et al., 2004).

A moderate spatial dependence (25<C0/(C+C0)<75%) were detected for IR, OM and co-variable of this properties, while silt and co-variable of IR and silt had the strong spatial dependence (C0/(C+C0)<25%). This results indicate that nugget effect were higher at IR and OM as compared to silt content. The moderate spatial dependence with high nugget effect for IR and OM might be hint to the tillage effect and cultivation. Cambardella and Karlen (1999) reported that spatially dependent may be controlled by intrinsic variations in soil characteristics such as mineralogy, and extrinsic variation such as tillage. In the study that conducted by Deurer et al. (2003), the macro pore networks are the primary effect of heterogeneity of the infiltration rate, a decreased macropore density at the lower depths might also result in decreased variance and increased spatial dependence of the field measurement IR.

A range of spatial variability indicated the distance beyond of semivariance become constant and the soil samples can be assumed to be spatially independent. In other hand, within the range, the measurement of the variable are correlated with each other. The range is important in term of controlling upper limits of the spatial dependencies prediction processes. The resulting semivariograms indicated a range of about 24 m for IR and range of about 52 and 58 m for OM and silt content, respectively. The range of cross-semivariograms differed with the level of auto-correlation variables were used. The cross-semivariogram of IR against OM showed that IR was correlated with OM up to the range of 36 m within they were moderately spatial dependent, while, the cross-semivariogram of IR against silt revealed a strong spatial dependence up to the range of 38 m. Generally, soil properties which are sensitive to management practices have a shorter geostatistical range (Ozgoz, 2007). Also Tesgaye and Hill (1998) reported that lower range could be due to a much sampling interval of 1 m in a relatively small area. Vieira et al. (1981) found a range of 50 m for 1280 field measured IR values at a field scale. A study
conducted by Cemek et al. (2007), and Sobieraj et al. (2004) also revealed spatial dependence in surface saturated hydraulic conductivity which is similar to IR, with range values of 17 and 25 m, respectively.

Kriging and cokriging procedures were used along semivariograms and cross-semivariograms to estimate IR values at unsampled points.

The cokriging procedure was applied to determine whether any advantage could be gained over kriging. The best auxiliary variable was determined using the cross validation. With cross validation, the prediction performance is checked by dropping actual data and estimating the properties of the location from the co-variables and neighboring data. Perfect cross validation agreement between true and predicted values would be reflected in having the lowest (near zero) mean absolute error (MAE) and general standard deviation (GSD) value.

This result showed that the MAE and GSD calculated from the cokriging with OM were slightly less than the MAE and GSD calculate from kriging method (Table 2), so indicated that the cokriging had no advantage over kriging when OM used as auxiliary variable. However using silt content as auxiliary data for the estimation of IR in cokriging method was consistently more effective than kriging to improve estimates of IR, and could reduce MAE by 15% as compared kriging method. many researcher have reported similar results to those above, for example Zhang et al. (1992) showed that with limited data, cokriging as compared with kriging, significantly improve estimation of particle size fraction in the areas of the field when using the reflectance of near infrared band as the auxiliary variable. Also Tarr et al. (2005) by means of cokriging method tested many auxiliary parameters, such as clay, soluble calcium, soluble magnesium, and depth of bed rock. They reported that clay content as auxiliary parameter was more suitable for surveying soil salinity in cokriging method. Finally, this study illustrated that the moderate spatial relationship between IR and silt content helped maintain the spatial information in IR at unsampled point.

4. Conclusion

Making successful design of the irrigation and drainage systems is one of the goals of the understanding the distribution of infiltration rate. This experiment illustrated the possibility of using kriging and cokriging methods. The obvious advantage of using such a predictive method arises from the fact that a large number of field measurements of infiltration rate are costly, time consuming, and cumbersome, whereas the method provides a means for predicting reliably the best estimate possible of the representative value from limited in situ measurement. The low relationship between IR and some soil properties can be attribute to the influence of other factor that were not able measured, such as tillage effect, root channels and cultivation. These factor causes infiltration rate change abruptly in space. Percentage of OM and silt were moderately correlated to IR within distances ranging from 36 to 42 m. The results of cross validation showed that cokriging no advantage over kriging
when OM use as auxiliary data. It was also conducted that silt content as an auxiliary data in the cokriging method is preferable to the OM content. This analysis can be applied in making decisions regarding agricultural and environmental land management.

References