Geostatistical-based index for spatial variability in soil properties

Edemar Appel Neto(1), Enio Junior Seidel(2)* and Marcelo Silva de Oliveira(1)

(1) Universidade Federal de Lavras, Departamento de Estatística, Lavras, Minas Gerais, Brasil.
(2) Universidade Federal de Santa Maria, Departamento de Estatística, Santa Maria, Rio Grande do Sul, Brasil.

ABSTRACT: The assessment of spatial variability of environmental variables such as soil properties is important for site-specific management. A geostatistical index that allows quantifying and characterizing the structure of spatial variability is fundamental in this context. Thus, this study aimed to develop a new spatial dependency index, called the Spatial Dependence Measure (SDM) for the spherical, exponential, Gaussian, cubic, pentaspherical, and wave semivariogram models; and comparing it with some of the indexes available in the literature. The SDM is also dimensionless, in the same way as the Spatial Dependence Index (SDI), also considering more parameters of the semivariogram, when compared to the Spatial Dependence Degree (SPD) and Relative Nugget Effect (NE) indexes. In a simulation data study, it is observed that the SDI and SDM indexes showed an advantage over the SPD (or NE). To exemplify the application of the SDM in the proposal for the classification of soil properties, we used estimates of geostatistical parameters presented in the two studies. The results indicate that the SDM can be a measure that, analyzed together with the SDI, can help to improve the description of the spatial variability structure. Thus, this study expands the number of geostatistical-based measures and increases the power of decision on the description of the degree of spatial variability of agricultural and soil attributes.

Keywords: within-field variability, spatial dependence, autocorrelation, semivariogram.
INTRODUCTION

Several studies have highlighted the importance of measuring the spatial variability of agricultural and soil properties using geostatistical indexes (Seidel and Oliveira, 2014, 2016; Appel Neto et al., 2018; Santos et al., 2018; Amaral and Della Justina, 2019; Leroux and Tisseyre, 2019). Such indexes are useful to assess the quality of the model fit on the semivariogram (Pazini et al., 2015; Oldoni and Bassoi, 2016; Büttow et al., 2017) and, consequently, to indicate whether kriging interpolation results in good quality maps (Appel Neto et al., 2018). In addition, these indexes allow the comparison of within-field variability in different research situations such as when comparing the variability of different types of soils, chemical or physical properties, different crops, among others.

One of the most common indexes in the Soil Science literature in Brazil to calculate the degree of spatial variability is presented in Cambardella et al. (1994), called the Relative Nugget Effect (NE), relating the nugget effect and the sill parameters. Another existing index is that proposed by Biondi et al. (1994), called the Spatial Dependence Degree (SPD), relating the contribution and the sill parameters.

An alternative for the inclusion of the range parameter into a measure of spatial dependence is the use of integral scales J1 and J2 (Russo and Jury, 1987). Han et al. (1994) presented a closed form for the integral scale J1, called Mean Correlation Distance (MCD), considering the contribution, sill, and range parameters. Despite considering more aspects of the semivariogram, integral scales are given in a unit of distance measurement as meters (m) or kilometers (km), which makes it difficult to propose a categorization to classify spatial dependence.

A Spatial Dependence Index (SDI) was proposed by Seidel and Oliveira (2014, 2016) and Appel Neto et al. (2018), which is a dimensionless index (free of measurement units), inspired by integral scale J1 (Russo and Jury, 1987) and based on spatial correlation area. This SDI index takes a specific form for each semivariogram model considered, being proposed for the spherical, exponential, Gaussian (Seidel and Oliveira, 2014, 2016), cubic, pentaspherical, and wave models (Appel Neto et al., 2018). For its construction, the SDI index considers the following parameters of the semivariogram: contribution, sill, range, and half of the maximum distance between sampled points. The spherical, exponential, Gaussian, cubic, pentaspherical, and wave semivariogram models were studied; according to Olea (2006), they are the most used by researchers.

In the matter of geostatistical indexes for measuring spatial variability, there is still room to expand this knowledge. Thus, following the same idea as Seidel and Oliveira (2014, 2016) and Appel Neto et al. (2018) in the construction of the SDI, the aim of this study was to develop a new spatial variability index, called the Spatial Dependence Measure (SDM), inspired by integral scale J2 (Russo and Jury, 1987), and comparing it with some of the indexes already exist in the literature.

MATERIALS AND METHODS

Equations 1 and 2 show the measurements NE (Cambardella et al., 1994) and SPD (Biondi et al., 1994), respectively.

\[
NE = \frac{C_0}{C_0 + C_1} \times 100 \tag{Eq. 1}
\]

\[
SPD = \frac{C_1}{C_0 + C_1} \times 100 \tag{Eq. 2}
\]
in which $C_0$ is the nugget effect; $C_1$ is the contribution; $C_0 + C_1$ is the sill; NE and SPD are complementary in the sense that SPD = 100 - NE. Thus, it was decided to use only the SPD in the analyses.

Equations 3 to 8 show the expressions of the SDI index for the spherical, exponential, Gaussian (Seidel and Oliveira, 2016), cubic, pentaspherical, and wave models (Appel Neto et al., 2018), respectively.

\[
SDI_{\text{spherical}} = 0.375 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 3

\[
SDI_{\text{exponential}} = 0.317 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 4

\[
SDI_{\text{gaussian}} = 0.504 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 5

\[
SDI_{\text{cubic}} = 0.365 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 6

\[
SDI_{\text{pentaspherical}} = 0.312 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 7

\[
SDI_{\text{wave}} = 0.589 \left( \frac{C_1}{C_0 + C_1} \right) \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100 
\]

Eq. 8

in which $C_1$ is the contribution; $C_0 + C_1$ is the sill; $a$ is the range; $MD$ is the maximum distance. The values 0.375, 0.317, 0.504, 0.365, 0.312, and 0.589 are the respective model factors ($MF$) of each of the semivariogram models. The min {} function is used to adjust the fact that the component $\frac{a}{0.5 \text{ MD}}$ is not necessarily limited in the amplitude from zero to one.

The SDM index was proposed following the same method as Seidel and Oliveira (2014, 2016) and Appel Neto et al. (2018). First, the spatial correlation measure (SCM) is calculated. In this case, what differentiates in the calculation of the SCM in relation to the integral scale $J_2$ is that in the SCM, the integration limit used is from zero until the value of the range ($a$). Equation 9 shows the spatial correlation measure obtained in general.

\[
SCM = \left\{ 2 \int_0^a \rho(h) h \, dh \right\} = \left\{ 2 \int_0^a \frac{(C_0 + C_1) - \gamma(h)}{C_0 + C_1} - h \, dh \right\} = MF \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{a}{0.5 \text{ MD}}} 
\]

Eq. 9

in which $\rho(h)$ is the correlogram function; $\gamma(h)$ is the semivariogram function; $h$ is the distance between sampled points; $C_0$ is the nugget effect; $C_1$ is the contribution; $C_0 + C_1$ is the sill; $a$ is the range; $MF$ is the model factor (specific to each semivariogram model).

Then, the SCM is multiplied by the inverse of half of the greatest distance between the georeferenced points in the sampling grid, according to equation 10. Half of the greatest distance between points was used in the same way as in Seidel and Oliveira (2016) and Appel Neto et al. (2018).

\[
SCM \left( \frac{1}{0.5 \text{ MD}} \right) = MF \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{a}{0.5 \text{ MD}}} \frac{1}{0.5 \text{ MD}} 
\]

Eq. 10

in which $C_1$ is the contribution; $C_0 + C_1$ is the sill; $a$ is the range; $0.5 \text{ MD}$ is half of the maximum distance between sampled points; $MF$ is the model factor (specific to each semivariogram model).
Then, by rearranging the equation 10, the Spatial Dependence Measure (SDM) is obtained. Equations 11 to 16 show the SDM index for the spherical, exponential, Gaussian, cubic, pentaspherical, and wave models, respectively.

\[
SDI_{\text{spherical}} = 0.447 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 11}
\]

\[
SDI_{\text{exponential}} = 0.422 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 12}
\]

\[
SDI_{\text{gaussian}} = 0.563 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 13}
\]

\[
SDI_{\text{cubic}} = 0.408 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 14}
\]

\[
SDI_{\text{pentaspherical}} = 0.378 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 15}
\]

\[
SDI_{\text{wave}} = 0.637 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{a}} \min \left\{ 1, \left( \frac{a}{0.5 \cdot MD} \right) \right\} 100 \quad \text{Eq. 16}
\]

in which \( C_1 \) is the contribution; \( C_0 + C_1 \) is the sill; \( a \) is the range; \( MD \) is the maximum distance. The values 0.447, 0.422, 0.563, 0.408, 0.378, and 0.637 are the respective model factors (MF) of each of the semivariogram models. The \( \min \{ \} \) function is used to adjust the fact that the component \( \frac{a}{0.5 \cdot MD} \) is not necessarily limited in the amplitude from zero to one.

To carry out the classification of the SDM, a methodology adapted from Seidel and Oliveira (2016) and Appel Neto et al. (2018) was used. Values were generated for the components \( \left( \frac{C_1}{C_0 + C_1} \right) \) and \( \left( \frac{a}{0.5 \cdot MD} \right) \), from 0.05 to 1.00, varying by 0.05. Afterward, these values were combined by multiplications \( \left( \frac{C_1}{C_0 + C_1} \right) \left( \frac{a}{0.5 \cdot MD} \right) \), which generated a vector of 400 values that was increased by a zero value. Thus, the vector of 401 values between 0 and 1 was multiplied by MF100 for each model. Finally, the median and the third quartile of each vector (one for each semivariogram model) were considered as cuts to categorize the SDM and classify the spatial dependence as weak, moderate, and strong, as shown in table 1.

To evaluate the performance of the SDM and compare it with some of the indexes used in the literature, 25 scenarios of spatial variability were simulated in the geoR package.
(Ribeiro Junior and Diggle, 2001), from R software (R Development Core Team, 2018). The scenarios were composed with the following parameters: contribution values of 10, 25, 50, 75, and 90 % of sill = 50; range values of 10, 25, 50, 75, and 90 % of 0.5MD = 70.71 m, in at 100 × 100 m sampling grid. Each scenario was replicated 100 times to mitigate possible variations in the simulation algorithm. With the 100 values generated from each scenario, the SPD, SDI, SDM and, as performance measures, the Mean Squared Error (MSE) and the Kriging Variance (KV) generated by cross-validation were calculated. Finally, Pearson’s correlations between the indexes and performance measures were calculated. In the case of a high spatial dependence structure, there are higher values in the SPD, SDI, and SDM indexes and lower values for the MSE and KV, so those negative correlations between the indexes and performance measures are expected. The results of the correlations are shown in table 2.

To exemplify the application of the SDM in the proposal for classification of soil properties, we use estimates of geostatistical parameters presented in the studies by Oldoni and Bassoi (2016) and Guedes et al. (2020). The authors used the SPD (or NE) and SDI indexes. From this, we also calculate the SDM index and apply its proposed classification. These results are shown in table 3.

### RESULTS AND DISCUSSION

The SDM differs from the SDI in the values of the model factor (MF) and also by the inclusion of a square root in the component $\frac{C_1}{C_0 + C_1}$ to give more weight to this component in the measurement of spatial variability. In addition, the SDM is also dimensionless (does not depend on units of measurement), in the same way as the SDI, also considering more parameters of the semivariogram, when compared to the NE and SPD indexes. The range parameter allows the evaluation of the spatial variability in the horizontal direction of the semivariogram, and the contribution and the sill parameters allow the evaluation of the spatial variability in the vertical direction of the semivariogram (Santos et al., 2018). The SDI and SDM indexes capture these aspects well in the whole area of the semivariogram graph, in other words, in both directions (horizontal and vertical) within the graph area, that is, in the direction of the horizontal axis of the semivariogram graph, and in the direction of the vertical axis of the semivariogram graph.

In general, in table 2, of the ten possible correlations between each index and performance measures, for the SPD there are six of them as strong, for the SDI there are eight of them as strong, and for the SDM also there are eight of them as strong. Thus, it is observed that the SDI and SDM indexes showed an advantage over the SPD. In addition, SDI was slightly better than SDM. Seidel and Oliveira (2014) found that the SDI had a slight advantage over the SPD, with a higher frequency of good correlations with the mean squared error of cross-validation. However, Amaral and Della Justina (2019) observed that the NE (or SPD) and SDI indexes did not perform as well as cross-validation measures in assessing the quality of kriging maps. In the sense of the relationship between the

### Table 2. Correlations between the Spatial Dependence Degree (SPD), Spatial Dependence Index (SDI), and Spatial Dependence Measure (SDM), and between Mean Squared Error (MSE) and Kriging Variance (KV) of the cross-validation, in semivariogram models

<table>
<thead>
<tr>
<th>Indexes</th>
<th>Spherical</th>
<th>Exponential</th>
<th>Gaussian</th>
<th>Cubic</th>
<th>Wave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>KV</td>
<td>MSE</td>
<td>KV</td>
<td>MSE</td>
</tr>
<tr>
<td>SPD</td>
<td>-0.67</td>
<td>-0.44</td>
<td>-0.92</td>
<td>-0.70</td>
<td>-0.79</td>
</tr>
<tr>
<td>SDI</td>
<td>-0.84</td>
<td>-0.97</td>
<td>-0.91</td>
<td>-0.97</td>
<td>-0.79</td>
</tr>
<tr>
<td>SDM</td>
<td>-0.80</td>
<td>-0.94</td>
<td>-0.90</td>
<td>-0.95</td>
<td>-0.72</td>
</tr>
</tbody>
</table>

Strong correlations: -1.00 to -0.70; moderate correlations: -0.69 to -0.30. The pentaspherical model is not implemented in the geoR package (Ribeiro Junior and Diggle, 2001) so that the simulation was not performed.
indexes and the range parameter, Santos et al. (2018) found that SDI had a moderate to a strong positive correlation with range parameter; and, the SPD had a weak to moderate negative correlation with range parameter.

Some researchers use metrics together to better assess spatial variability and optimize decision-making: Taylor et al. (2007) using the NE and the MCD; Souza et al. (2008) using the NE and the integral scale J2; Oldoni and Bassoi (2016) using SPD and SDI; Amaral and Della Justina (2019) and Guedes et al. (2020) using the NE and SDI. These findings show the need for further studies to propose and evaluate the performance of the indexes, mainly for the SDM that is being proposed.

From table 3, it can be seen that the SPD index generates four strong classifications, ten moderate classifications and one weak. The SDI generates five strong classifications, seven moderate, and three weak. While the SDM index generates four strong ratings, four moderate, and seven weak. Considering a joint assessment of the three indexes, it is possible to verify that sand, available water, soil density, 60 days after pruning (DAP), and 63 DAP were classified as strong in at least two of the indices. The properties silt, clay, 100-101 DAP, 78 DAP, calcium, and pH were classified as moderate in at least two of the indexes. The properties 57 DAP, carbon, and magnesium were classified as weak by at least two of the indexes. However, the 91 DAP property had the three discordant classifications.

**CONCLUSIONS**

The Spatial Dependence Measure (SDM) can be a measure that, analyzed together with the Spatial Dependence Index (SDI), can help to improve the description and classification of the spatial variability structure. Thus, this study expands the number of geostatistical-based measures and increases the power of decision on the description of the degree of spatial variability of agricultural and soil properties.
ACKNOWLEDGEMENTS

The authors would like to thank FAPEMIG and CNPq for their support in carrying out this research.

AUTHOR CONTRIBUTIONS

Conceptualization: 🅗 Enio Júnior Seidel (equal) and 🅗 Marcelo Silva de Oliveira (equal).
Methodology: 🅗 Edemar Appel Neto (equal), 🅗 Enio Júnior Seidel (equal), and 🅗 Marcelo Silva de Oliveira (equal).
Formal analysis: 🅗 Edemar Appel Neto (lead), 🅗 Enio Júnior Seidel (supporting), and 🅗 Marcelo Silva de Oliveira (supporting).
Writing - original draft: 🅗 Edemar Appel Neto (equal) and 🅗 Enio Júnior Seidel (equal).
Writing - review and editing: 🅗 Edemar Appel Neto (equal), 🅗 Enio Júnior Seidel (equal), and 🅗 Marcelo Silva de Oliveira (equal).
Supervision: 🅗 Enio Júnior Seidel (equal) and 🅗 Marcelo Silva de Oliveira (equal).
Funding acquisition: 🅗 Edemar Appel Neto (equal) and 🅗 Marcelo Silva de Oliveira (equal).

REFERENCES


Errata RBCS 2020-0086

In the article “Geostatistical-based index for spatial variability in soil properties” [Rev Bras Cienc Solo. 2020;44: e0200086. DOI: 10.36783/18069657rbcs20200086], on page 4 (Equations 11 to 16), where it is presented:

\[
SDI_{\text{spherical}} = 0.447 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 11}

\[
SDI_{\text{exponential}} = 0.422 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 12}

\[
SDI_{\text{gaussian}} = 0.563 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 13}

\[
SDI_{\text{cubic}} = 0.408 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 14}

\[
SDI_{\text{pentaspherical}} = 0.378 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 15}

\[
SDI_{\text{wave}} = 0.637 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 16}

**It should be presented:**

\[
SDM_{\text{spherical}} = 0.447 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 11}

\[
SDM_{\text{exponential}} = 0.422 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 12}

\[
SDM_{\text{gaussian}} = 0.563 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 13}

\[
SDM_{\text{cubic}} = 0.408 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 14}

\[
SDM_{\text{pentaspherical}} = 0.378 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 15}

\[
SDM_{\text{wave}} = 0.637 \left( \frac{C_1}{C_0 + C_1} \right)^{\frac{1}{4}} \min \left\{ 1; \left( \frac{a}{0.5 \text{ MD}} \right) \right\} 100
\]
\text{Eq. 16}