A contribution to inverse-distant method as a step towards a cartography’s research freedom
N. Protic and S. Protic
Institut for Soil Science, T.Drajzera 7, Belgrade, YU E-mail: soilscis@eunet.yu
S.R.Protronik, Njegoseva 44, Belgrade, YU
E-mail: protron@eunet.yu

Abstract

A contribution to inverse-distant method is presented with a corresponding software for the calculation and cartographic presentation of results. The aim is to obtain the tool by which a researcher can express his creation in spatial interpretation with an adequate control of results.

The method is based on inverse-distant prediction of the value of grid data by the application of the following basic function: \( y = y_1 e^{-x^2} \). The basic function is extended by the introduction of influencing coefficients: global, which affects the “global” form of the function, class, resulting from any given number of classes, i.e. the class to which the true values belong, and grid coefficient whose value depends on grid density. In addition, in the computation, it is also possible to choose the boundaries of influence between data pairs, as well as the number of elements in the grid. In cartographic presentation, the user can determine the zone to be cut out, i.e. part of the field without a sufficient number of measured data, so that it is not necessary to create a special blank file.

By introducing the modifying elements to the basic equation, the user can select the aims of cartographic presentation of his results, with simultaneous light control of reliability of the obtained grid values.
1 Introduction

Soil science uses a great number of softwares based on theoretical foundations and methods that combine the functions of spatial analysis and the corresponding cartographic presentations (See all References). In the majority of currently used methods, spatial assessment is based on objective functions (variograms or other functions), which considering the number and quality of measurements, as well as the (non)stationarity of the variable in the researched range, ensure the minimum estimation error in nonmeasured value compared to true value. After the system of estimation has been optimized, the function of presentation and control of efficiency is achieved by simple mathematical operations. Of course, such an approach is irreplaceable if we, as spatial researchers, are satisfied with fixed rules of statistical-mathematical models and with the exactness of such solutions. However, in the assessment of environmental quality, we are often dissatisfied with an exact cartographic presentation resulting from “objective” functions. On the contrary, we are often obliged to regard spatial analysis in the context of potential future change intensities. For this purpose, the developed method offers a choice of spatial presentations, which is optimal from the aspect of a specific requirement, with simultaneous control of efficiency, considering the defined boundary conditions.

2 Material and method

Similar to other methods of map analysis, this method is also developed on the principles of estimation of nonmeasured values based on true values. Namely, if in the two-dimensional field with known co-ordinates, some points have true values, this method makes it possible to obtain the values of all points in the field, with probability of exactness according to the function.

The method is based on the hypothesis that the exactness of true values is maximal (100%) and that the probability of values of the points surrounding the measured point in all directions decreases according to the function (Fig. 1). This function is taken as the most adequate mathematical description of natural forms (mountain of volcanic origin, sand flown in the sand-glass, etc.), with the main hypothesis on the probability that, in close vicinity of a true value, it decreases very little and then, as the distance increases, it is increasingly greater.
At greater distances, the hypothesized value asymptotically approaches zero, because the probability that the value of the point is equal to the true value is increasingly lower. The basic equation which describes this change has the following form:

\[ y = y_t \cdot e^{-x^2} \]  

[1]

\( e \) the base of natural logarithm  
\( x \) distance from the true value  
\( y_t \) true value

The final equation has three influencing coefficients: global, class, and grid coefficients. The values of global and class coefficients determine whether the curve is narrow (curve 1, Fig. 1) or wide (curve 2, etc.), i.e. whether the observed true value more or less influences the surrounding points in the grid. The global influencing coefficient is arbitrary (0.001-1) and it affects the global “weight” of all surrounding points in the grid, while class coefficient results from the class which the true value belongs to. The class is determined by the number of true points of similar values.

Example:
True values of fluorine (Fl) are presented in Table 1.
Table 1. Point coordinates and values

<table>
<thead>
<tr>
<th>Nr.</th>
<th>x(km)</th>
<th>y(km)</th>
<th>F1(mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.2</td>
<td>11.3</td>
<td>14.20</td>
</tr>
<tr>
<td>2</td>
<td>50.4</td>
<td>29.9</td>
<td>30.15</td>
</tr>
<tr>
<td>3</td>
<td>71.6</td>
<td>17.8</td>
<td>23.82</td>
</tr>
<tr>
<td>4</td>
<td>23.2</td>
<td>67.3</td>
<td>19.32</td>
</tr>
<tr>
<td>5</td>
<td>27.6</td>
<td>22.8</td>
<td>31.44</td>
</tr>
<tr>
<td>6</td>
<td>15.7</td>
<td>82.6</td>
<td>17.92</td>
</tr>
<tr>
<td>7</td>
<td>83.8</td>
<td>40.1</td>
<td>27.37</td>
</tr>
<tr>
<td>8</td>
<td>72.1</td>
<td>71.5</td>
<td>16.23</td>
</tr>
<tr>
<td>9</td>
<td>55.4</td>
<td>57.4</td>
<td>28.91</td>
</tr>
</tbody>
</table>

The program starts by finding the minimal and maximal values of columns x, y and Fl (x_{min}, x_{max}, y_{min}, y_{max}, Fl_{min} and Fl_{max}).

The obtained values are as follows:

\[ x_{min}=10.2, \quad x_{max}=83.8, \quad y_{min}=11.3, \quad y_{max}=82.6, \quad Fl_{min}=14.20, \quad Fl_{max}=31.44 \]

The observed field is plotted so that its lower left corner is determined by the coordinate \((x_{min}, y_{min})\), and its upper right corner, by the coordinate \((x_{max}, y_{max})\). The obtained field is divided both by x and by y-axes by the predetermined number of segments (grid density) and in our example it is \(NrSeg=15\) (Fig. 2). In this way the sizes of grid segments are determined for both axes:

\[
\Delta x = \frac{x_{max} - x_{min}}{NrSeg} = \frac{83.8 - 10.2}{15} = 4.907 [km]
\]

\[
\Delta y = \frac{y_{max} - y_{min}}{NrSeg} = \frac{82.6 - 11.3}{15} = 4.753 [km]
\]
2.1. Class coefficient

This phase determines the classes of each of the true values of fluorine (Fl). The number of classes is arbitrary and in our case it is assumed to be \( NrCl = 3 \). This datum segments the difference between \( Fl_{\text{max}} \) and \( Fl_{\text{min}} \) into three equal parts, and then the data are classified according to their values. In our example, the class width is:

\[
ClW = \frac{Fl_{\text{max}} - Fl_{\text{min}}}{NrCl} = \frac{31.44 - 14.20}{3} = 5.75 \text{[mg/kg]}
\]

The classified data are presented in Table 2.

In this way, we introduce the intensity of influence of the value of a particular point on its surroundings, depending on its class. If the number of classes is greater, the number of data per class decreases, which logically reduces the value of the coefficient, but simultaneously reduces also the class segment.
Finally, class coefficient (bottom row, Table 2), is expressed as the normalized value of the number of data in a particular class as per the total number of data, which is in our example NV=9.

<table>
<thead>
<tr>
<th>Class Nr.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Seg.</td>
<td>Fl(14.20-19.95)</td>
<td>Fl(19.95-25.70)</td>
<td>Fl(25.70-31.44)</td>
</tr>
<tr>
<td>Or.Nr.of val.</td>
<td>1, 4, 6, 8</td>
<td>3</td>
<td>2, 5, 7, 9</td>
</tr>
<tr>
<td>NVC</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>(k_{ci} = \frac{NVC}{NV})</td>
<td>(\frac{4}{9}=0.444)</td>
<td>(\frac{1}{9}=0.111)</td>
<td>(\frac{4}{9}=0.444)</td>
</tr>
</tbody>
</table>

NVC: number of data in a class  
NV: number of true values  
k_{ci}: class coefficient

### 2.2. Grid coefficient

Under a lower number of segments, the grid becomes thinner, so the points of true values are more likely to enter one of the fields in the grid. If the number of segments is higher, the probability decreases. This phenomenon is introduced in the calculation as grid coefficient:

\[
k_{Gr} = \frac{NV}{NrSeg^2}
\]

It is the quotient of the number of true values and the overall number of segments in the grid.

### 2.3. Estimation of reliability of grid values – probability matrix

The modified equation for the estimation of the of grid data is:
Development and Application of Computer Techniques to Environmental Studies

\[ V_{tg} = \sum_{i=1}^{n} e^{\frac{\left( \frac{R_i}{R_{\text{max}}} \right)^2}{k_{Gl} \cdot k_{Cl(i)} \cdot k_{Gr}}} \]  

\[ V_{tg} \] cumulative probability of exactness of grid data
\[ e \] the base of natural logarithm
\[ R_i \] distance between the observed point in the grid and the \( i^{\text{th}} \) true value
\[ R_{\text{max}} \] maximal distance between a point in the grid and true points
\[ k_{Gl} \] preselected global coefficient
\[ k_{Cl(i)} \] class coefficient of \( i^{\text{th}} \) point of true value
\[ k_{Gr} \] grid coefficient

Each point in the grid is connected with all points with true values. The procedure is shown by the point “A” (Fig. 3). Based on known coordinates (\( x, y \)) of the true value “1” and the known coordinate of the point “A”, the influencing radius of point “1” to point “A” (\( R_1 \)) is calculated by Pythagoras’ theorem.

\[ R_1 = \sqrt{(x_A - x_1)^2 + (y_A - y_1)^2} \]

Coordinates of point “A”:
\[ x_A = x_{\text{min}} + 6 \cdot \Delta x = 10.2 + 6 \cdot 4.907 = 39.642 [\text{km}] \]
\[ y_A = y_{\text{min}} + 3 \cdot \Delta y = 11.3 + 3 \cdot 4.753 = 25.559 [\text{km}] \]

The procedure is repeated for other points with true values (Table 3).
Figure 3. Probability estimation for a grid point

\[ R_{\text{max}} \text{ is determined from the obtained values of the influencing radius:} \]

\[ R_{\text{max}} = R_6 = 61.862 \text{[km]} \]

By applying the pre-selected global coefficient, computed class coefficients (1 to 9, Table 2), grid coefficient and influencing radius (Table 3), we can estimate the value of cumulative probability of exactness of grid point “A” data.

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.713</td>
<td>11.600</td>
<td>32.886</td>
<td>44.863</td>
<td>12.354</td>
</tr>
</tbody>
</table>

In this place, each calculated value of probability is compared to previously not referred to and pre-selected boundary of influence \( \text{GrUt} \in (\geq 0.01 \leq (e^{-1} = 0.37)) \), arbitrarily determined by the user. If for an influencing radius \( R_i \), \( V_{tg}(i) < \text{GrUt} \), the value of \( V_{tg}(i) \) is reduced to zero, it is not included in the overall sum of cumulative probability (Fig. 4).
For the pre-selected values:

- $k_{gi} = 0.1$ global coefficient
- $GrUt = 0.01$ boundary of influence

the values of the probabilities per influencing radiiuses are given in Table 4.

The value of cumulative probability for point “A” is:

$$V_{tg}[A] = \sum_{i=1}^{9} V_{tg}(i) = 0.8603$$

<table>
<thead>
<tr>
<th></th>
<th>$V_{tg}(1)$</th>
<th>$V_{tg}(2)$</th>
<th>$V_{tg}(3)$</th>
<th>$V_{tg}(4)$</th>
<th>$V_{tg}(5)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>without $GrUt$</td>
<td>0.0018</td>
<td>0.4530</td>
<td>0.0000</td>
<td>0.4073</td>
<td>0.0000</td>
</tr>
<tr>
<td>with $GrUt$</td>
<td>0</td>
<td>0.4530</td>
<td>0</td>
<td>0.4073</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$V_{tg}(6)$</th>
<th>$V_{tg}(7)$</th>
<th>$V_{tg}(8)$</th>
<th>$V_{tg}(9)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>without $GrUt$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0006</td>
</tr>
<tr>
<td>with $GrUt$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The procedure is repeated for all grid points and the probability matrix is formed. The value of cumulative probability $V_{tg}(i)$ for the particular grid points can be higher than one. For this reason, normalization is performed by means of the maximal value $V_{tg\text{max}}$ in the probability matrix:

$$V_{nor}[i] = \frac{V_{tg}(i)}{V_{tg\text{max}}}$$  \[4\]
2.4. Estimation of value matrix

Value matrix is determined parallel with probability matrix. The minimal influencing radius is determined simultaneously with the maximal radius (Table 3):

\[ R_{\text{min}} = R_2 = 11.600 \text{[km]} \]

Grid point “A” obtains the value of the nearest true point, and in this case it is point “2” (Table 1). Its value is:

\[ V_r[A] = F_l_2 = 30.15 \text{[mg / kg]} \]

In the same way, the values of other grid points are estimated and the value matrix is formed.

2.5. Blank zone

In the processing of the assumed values of grid points, there might be field parts without a sufficient number of true data, so they should be cut out. The user decides on the distance from true points to which the non-measured grid points will be valued. It is normal that the value of this radius is greater than the diagonal of a grid field.

\[ R_{CZ} \geq \sqrt{\Delta x^2 + \Delta y^2} \]

In our example, the zone was not pre-selected, so the radius of cutting is equal to the diagonal of the entire field.

2.6. Final calculation of grid point values

The value of point “A” in our example is that which belongs to the nearest true value. This hypothesis is logical, but the value must be reduced to the value corresponding to probability of exactness:

\[ F_{l_p}[A] = V_{nor}[A] \cdot V_r[A] \]

\[ [5] \]

\( F_{l_p}[A] \) assumed value of Fl in grid point “A”
3 Results

By applying the described procedure to the remaining grid points, we can plot the graphical presentation of probability and the values in 2D and 3D, Figures 5 - 8, which simultaneously represent the final results.

By selecting different values of the global coefficient, class coefficient and grid coefficient, it is possible to get a new spatial interpretation for the same grid points, with a simultaneous control of the model, expressed through probability matrix.

4 Conclusions

The proposed software model enables the user to express his specific requirements, by the modules of the basic equation with the influencing coefficients. As the final solutions can be planned, by successive analyses it is possible to obtain different levels of map interpretativeness. Taking into account the density of grid data and the planned objectives, the desired reliability of the cartographic presentation can be readily and simultaneously controlled.
Development and Application of Computer Techniques to Environmental Studies

Figure 5. Probability of data exactness (2D)

Figure 6. Probability of data exactness (3D)
Figure 7. Assumed values of Fl (2D)

Figure 8. Assumed values of Fl (3D)
5 References

13. Protic, N., Jelena Moskovljievic, Vesna Mrvic, Branka Brebanovic, and A. Dinovic: Possibility of formalizing the procedure of
Development and Application of Computer Techniques to Environmental Studies


