Comparison of different types of Factorial Kriging maps in an environmental case study

Batista, A. C.¹, Ferreira da Silva, E. ¹, Sousa, A. J. ² and Cardoso Fonseca, E. ¹

¹ Departamento de Geociências, Universidade de Aveiro, Campus de Santiago, 3810-190 Aveiro, 2 CVRM / Centro de Geo-sistemas, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL

Key words: Factorial kriging, external drift, environmental geochemistry

Abstract

The main objective of this paper is to study the influence of different external drifts in the Factorial Kriging maps, and to find out if the type of external drift influences the resulting maps.

This study was conducted in a 336 km² area, covering the Águeda municipality (Central Portugal). This municipality is characterised by high east west geological and land use asymmetry. The west side has a considerably higher geological and land use variability. The main settlements and industries are located in this side. However hydrologically this asymmetry doesn’t exist. Taking into account that information, nine elements were analysed on a 1095 stream sediment samples set by atomic absorption: Cu, Pb, Zn, Ni, and Cr mainly related with anthropogenic activities, and Mn, Fe and Cd mainly related with geology.

An initial statistical study was undertaken involving the determination of the basic statistical parameters, as well as a PCA analysis, in order to identify the main relations between variables. The variographic study, upon which the Factorial Kriging was based, included the adjustment of the theoretical variograms to the experimental variograms by a semi-automatic process, limited to the estimation of parameters such as sills and ranges chosen by the authors. This variographic study was done on the normalised data set.

From this study three types of Factorial kriging maps were produced: 1- from the classical Factorial Kriging, 2- Factorial Kriging with external drift – where the drift is

¹ E-mail: constanca@geo.ua.pt, Fax: +351 234 401559
related with the lithological information (the sample co-ordinates on the second axis resulting from the PCA analysis) and 3- Factorial Kriging with external drift - where the drift is based on the sample codification according to the land use. These maps were compared with the kriging map of the initial values of each variable.

The main conclusions that rise from this study are: 1- the introduction of secondary information, in the Factorial Kriging maps, allows a better definition of the anomalous areas; 2- the type of external drift influences the resulting maps, since the external drift has to be highly correlated with the variables that are being studied.

1. Introduction

Factorial Kriging (Sandjivy, 1984; Wackernagel, 1995) has been used in several areas by several authors (Sousa, 1988, 1994; Wackernagel and Sanguinetti, 1993; Jimenez-Espinosa et al.; 1993, Barata et al, 1997, Wen et al, 1997; Nunes et al, 2000, Batista et al, 2001a), b) as it is an important technique of multivariate estimation. The characterisation of each of the structures (given by the variogram) in which each regionalized variable, at several scales, can be decomposed is done by Factorial Kriging, leading to different types of mapping (Batista, 1998; Batista et al, 2001b))

In order to improve the cartography of the different spatial components of a regionalized variable (primary variable), we can introduced the information given by an external, complementary variable (secondary variable) (Wackernagel, 1995), as long as: (a) the variables are correlated and (b) the second variable is known in all the domain.

Taking all this information in consideration a new methodology to estimate the spatial components was created (Batista, 1998, Batista et al, 2001b)): the Factorial Kriging with External Drift.

The Factorial Kriging system was modified in order that the estimation of the spatial components includes the same relation between the two variables (primary and second variable). The estimation variance has therefor to be minimised, for each structure \( u \), taking into consideration two restrictions: 

\[ \sum_{i=1}^{n} \lambda_i^u = 0 \]
(Wackernagel, 1995), since the spatial components have, by definition, null expectation and (2) \[ \sum_{i=1}^{n} \lambda_i^u \ s(x_i) = s(x_0) \] (Wackernagel, 1995, pag. 193).

This second restriction comes from the relation between the primary variable \( z(x_i) \) as a particular realisation of the random function) and secondary variable \( s(x_i) \):

\[
E\{Z(x_i)\} = a_0 + b_1 \ s(x_i).
\]

Given an external drift, the primary variable can be estimated by a linear combination of the available data:

\[
z^*(x_0) = \sum_{i=1}^{n} \lambda_i \ z(x_i)
\]

where the weights depend on the external drift function \( s(x_i) \), cf. (Wackernagel, 1995):

\[
\sum_{i=1}^{n} \lambda_i \ s(x_i) = s(x_0)
\]

and

\[
E\{Z^*(x_0)\} = a_0 + b_1 s(x_0)
\]

As a result it emerges the following kriging system (Batista et al, 2001b):

\[
\begin{align*}
\sum_{j=1}^{n} \lambda_j^u C(x_i - x_j) - \mu_1 - \mu_2 s(x_i) &= C^u(x_i - x_0), \text{ for } i = 1, \ldots, n \\
\sum_{j=1}^{n} \lambda_j^u &= 0 \\
\sum_{j=1}^{n} \lambda_j^u s(x_j) &= s(x_0)
\end{align*}
\]
where \( C^u(x_i - x_0) \) is the elementary covariance between the value on \( x_i \) sampling point and \( x_0 \) the point to be estimated and, \( N_s \) the number of spatial structures,

\[
C(x_i - x_0) = \sum_{u=0}^{N_s} C^u(x_i - x_0) .
\]

2. Case Study: Águeda Region

Águeda municipality, with an area nearly 334.3Km\(^2\), is located in the central-Northeast of Portugal, more specifically in the sub-area of the Low-Vouga. This municipality belongs to the district of Aveiro. This municipality is limited by the Caramulo Mountain to south and east, by the Vouga River valley and Arestal Mountain to north, and by the morphologic unit named Ria de Aveiro to west. The industry is very diversified and mainly related to electroplating, foundries, metal processing and manufacturing industries.

Some studies published in 1996 estimated that there are a total of 600 industries in this region. Nearly 20% of these are electroplating industries. In the east part of Águeda we have one abandoned Cu, Pb and Zn mine (Talhadas mine) with some problems connected to tailing deposits and acid mine drainage.

2.1 Geology

The geological mapping of the Águeda area was done by Ávila Martins (1962), Godinho (1974) and Mendia de Castro (1986). Figure 1 shows the simplified geological map of Águeda municipality where it is possible to clearly identify an east-west asymmetry.

The east part is characterised by the presence of schist from the “Complexo Xisto-Grauváquico”. The west part is mainly characterised by the presence of Triassic sandstone (Palain, 1976), Cretaceous sandstone (Barbosa, 1981), Plio-Plistocenic alluvium terraces (Mendia de Castro, 1986).
Figure 1 – Simplified geology map of the Águeda municipality. (1) – Alluvium; (2) – Sandstone; (3) – Cretaceous; (4) – Triassic; (5) – Permian; (6) – Ordovician; (7) – Schist (Complexo Xisto-Grauváquico); (8) – Talhadas Granite; (9) – Dykes and (10) – Faults.

2.2 Land Use

The land use codification was established by overlaying the sample distribution map to the Land Cover Corine map (scale 1:100 000) which was based on the Landsat 5 images (Fig. 2).

This municipality is also characterised by a high land use east – west asymmetry.

Figure 2 – Land use map of the Águeda municipality based on the Land Cover Corine map.
2.3 Sampling and Analysis

A total of 1095 stream sediment samples were collected in the area. The stream sediment samples were initially homogenised and oven-dried at 60°C for 24 hours. One portion of the dried sample was removed, sieved and the < 80 mesh fraction was retained for chemical analysis.

The Cu, Pb, Zn, Fe, Mn, Co, Ni and Cr concentrations were determined by flame atomic absorption spectrophotometry technique (AAS), following a hot HCl-HNO₃-HF attack (Lecomte & Sondag 1980).

The precision of the analytical results was estimated by replicate analysis (Garrett 1973). Analytical precision, defined as the percent relative variation at the 95% confidence level, ranged from 4 to 13%, for Cu, Pb, Zn, Fe, Mn, Co, Ni and Cr considering the P₀.05 significance level.

2.4 Results

2.4.1 Preliminary Analysis

The main statistical parameters for all variables are shown in table 1. The correlation coefficients between the initial values and between the logarithm values are shown in table 2. Almost all variables show a skewed distribution, with excess of low values. With the exception of Fe and Cd, all variables present very high variation coefficients.

<table>
<thead>
<tr>
<th>Variable:</th>
<th>Cu</th>
<th>Pb</th>
<th>Zn</th>
<th>Co</th>
<th>Ni</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>56</td>
<td>71</td>
<td>193</td>
<td>27</td>
<td>60</td>
<td>84</td>
<td>215</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>33</td>
<td>52</td>
<td>99</td>
<td>19</td>
<td>34</td>
<td>31</td>
<td>114</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Mode</td>
<td>29</td>
<td>40</td>
<td>79</td>
<td>15</td>
<td>24</td>
<td>34</td>
<td>60</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Variance</td>
<td>21192</td>
<td>17656</td>
<td>756189</td>
<td>1212</td>
<td>177860</td>
<td>2088208</td>
<td>128474</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>146</td>
<td>133</td>
<td>870</td>
<td>35</td>
<td>422</td>
<td>1445</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Minimum</td>
<td>3</td>
<td>12</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>0.23</td>
<td>0.6</td>
</tr>
<tr>
<td>Maximum</td>
<td>2788</td>
<td>2842</td>
<td>27112</td>
<td>662</td>
<td>13618</td>
<td>47730</td>
<td>4482</td>
<td>33.6</td>
<td>9.3</td>
</tr>
<tr>
<td>Variation Coeff.</td>
<td>262</td>
<td>188</td>
<td>451</td>
<td>128</td>
<td>697</td>
<td>1725</td>
<td>167</td>
<td>59</td>
<td>35</td>
</tr>
</tbody>
</table>
Table 2 - Correlation coefficients between the initial values and between the logarithm values

<table>
<thead>
<tr>
<th></th>
<th>Cu</th>
<th>Pb</th>
<th>Zn</th>
<th>Co</th>
<th>Ni</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>1.00</td>
<td>0.15</td>
<td>0.18</td>
<td>0.00</td>
<td>0.49</td>
<td>0.44</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.16</td>
</tr>
<tr>
<td>Pb</td>
<td>1.00</td>
<td>0.17</td>
<td>0.01</td>
<td>0.05</td>
<td>0.05</td>
<td>0.00</td>
<td>0.13</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>1.00</td>
<td>0.03</td>
<td>0.22</td>
<td>0.24</td>
<td>0.03</td>
<td>0.03</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co</td>
<td>1.00</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.61</td>
<td>0.21</td>
<td><strong>0.51</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>1.00</td>
<td><strong>0.98</strong></td>
<td>0.01</td>
<td>0.06</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td>1.00</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>1.00</td>
<td>0.21</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td>1.00</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cd</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Log Values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>1.00</td>
<td>0.36</td>
<td>0.45</td>
<td>0.24</td>
<td>0.46</td>
<td>0.4</td>
<td>0.21</td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>Pb</td>
<td>1.00</td>
<td><strong>0.58</strong></td>
<td>0.21</td>
<td>0.29</td>
<td>0.25</td>
<td>0.14</td>
<td>0.26</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>1.00</td>
<td><strong>0.63</strong></td>
<td>0.47</td>
<td>0.44</td>
<td>0.44</td>
<td><strong>0.52</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co</td>
<td>1.00</td>
<td><strong>0.58</strong></td>
<td>0.20</td>
<td>0.68</td>
<td>0.47</td>
<td><strong>0.73</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>1.00</td>
<td><strong>0.74</strong></td>
<td>0.36</td>
<td>0.41</td>
<td><strong>0.51</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td>1.00</td>
<td>0.05</td>
<td>0.33</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mn</td>
<td>1.00</td>
<td>0.33</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe</td>
<td>1.00</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cd</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is clear that significant correlations were found between the studied elements when there are calculated with the logarithmic values e.g. Zn with Pb, Co with Zn, Zn and Co with Ni, Cr with Ni, Mn with Co and Zn, Co, Ni with Cd. The results obtained indicate that those variables are logarithmic related.

Taking into account the information obtained by univariate and bivariate statistics Principal Component Analysis (PCA analysis) considering the logarithmic values (PCALog) was done. The main results are shown on Table 3. The first two components account for about 62% of the total explained variance. Figure 3 presents the variables and samples plots on the first factorial plan. All variables have positive co-ordinate on the first principal component (F1).

The second component (F2) separates two main groups: Mn – Co – Cd – Fe (mainly lithological), Zn – Ni and Pb – Cu – Cr (due to pollution).

![Figure 3 – First factorial plan (F1 versus F2): a) for the variables and b) for the samples.](image)
Table 3 – Eigen values and percentage of explained variance for the first 3 principal components. Scores of all variables in those components.

<table>
<thead>
<tr>
<th></th>
<th>F1 - Axis 1</th>
<th>F2 - Axis 2</th>
<th>F3 - Axis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>4.21</td>
<td>1.39</td>
<td>0.93</td>
</tr>
<tr>
<td>Percentage</td>
<td>46.76</td>
<td>15.45</td>
<td>10.44</td>
</tr>
<tr>
<td>Cu</td>
<td>0.57</td>
<td>-0.40</td>
<td>-0.17</td>
</tr>
<tr>
<td>Pb</td>
<td>0.54</td>
<td>-0.29</td>
<td>-0.69</td>
</tr>
<tr>
<td>Zn</td>
<td>0.83</td>
<td>-0.14</td>
<td>-0.23</td>
</tr>
<tr>
<td>Co</td>
<td>0.78</td>
<td>0.51</td>
<td>0.11</td>
</tr>
<tr>
<td>Ni</td>
<td>0.83</td>
<td>-0.23</td>
<td>0.35</td>
</tr>
<tr>
<td>Cr</td>
<td>0.59</td>
<td>-0.61</td>
<td>0.43</td>
</tr>
<tr>
<td>Mn</td>
<td>0.60</td>
<td>0.57</td>
<td>0.02</td>
</tr>
<tr>
<td>Fe</td>
<td>0.61</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>Cd</td>
<td>0.73</td>
<td>0.34</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

The PCALog results show that the second principal component best separates the main lithologies present in the area (after a 25° rotation of the samples co-ordinates).

2.4.2. Variography

A variographic study of all the variables was conducted in order to determine the variables spatial structure. From all analysed variables, only three of them are presented:

a) **Cu** and **Pb** - characterising anthropogenic pollution i) related with industrial activities and ii) related with the Talhadas mine, located in the NE part of the study area.

b) **Zn** – related to anthropogenic pollution either industrial either from urban activities.

The variogram parameters upon which the Factorial Kriging was based are shown, for the chosen variables, in Table 4. These variables have different directional spatial behaviour, since Cu and Pb have an anisotropic behaviour and Zn has an isotropic behaviour.

Those parameters were calculated with the standardised data (subtracted the average to each value and divided by the standard deviation).
A semi-automatic process was used for fitting the theoretical variograms to the experimental variograms, limited to the estimation of the sills and ranges chosen by the authors.

Table 4 – Variogram parameters for the chosen variables.

<table>
<thead>
<tr>
<th>Variogram parameters</th>
<th>$C_0$</th>
<th>$C_1$</th>
<th>$A_1$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu – Spherical Model</td>
<td>0.14</td>
<td>0.532</td>
<td>90° - 3360; 0° - 1680</td>
</tr>
<tr>
<td>Pb - Spherical Model</td>
<td>0.308</td>
<td>0.715</td>
<td>90° - 3360; 0° - 1680</td>
</tr>
<tr>
<td>Zn - Spherical Model</td>
<td>0.14</td>
<td>0.88</td>
<td>2160</td>
</tr>
</tbody>
</table>

* $C_0$ - is the nugget effect; $C_1$ - is the sill of the first structure and $A_1$ - is the range of the first structure.

2.4.3. Two Types of External Drift

By using different types of external drifts into the Factorial Kriging maps, we tried to evaluate the influence of the choice of the external drift.

For that purpose two external variables were chosen:

a) The sample’s co-ordinates on the second axis resulting from the PCA analysis (PCALog). These co-ordinates were calculated on the second axis after a 25° sample’s rotation (clock wise). This rotation to the co-ordinates was done since that way the separation of the main lithologies, by the second axis, was clearer.

b) The sample’s co-ordinates on a discriminant axis. The grouping variable for this discrimination was the land use codification. This codification was done by overlaying the sampling map and the Corine Land Cover map. The main land use codes are shown in table 5.

Table 5 – Main land use codes.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Number of samples</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>Urban Occupation</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Industrial Activities</td>
</tr>
<tr>
<td>3</td>
<td>437</td>
<td>Agricultural Land</td>
</tr>
<tr>
<td>4</td>
<td>636</td>
<td>Forest</td>
</tr>
</tbody>
</table>
2.4.4 Factorial Kriging

Figure 5, 6 and 7 show the Cu, Pb and Zn maps. The Factorial Kriging maps, corresponding to each of the spatial structures, were constructed, using the variogram parameters with the normalised data. For each variable the maps of the nugget effect ($C_0$) and of the first structure ($C_1$) are presented:

a) without external drift;

b) using the samples second axis (after a 25° rotation) as external drift;

c) using the samples co-ordinates on the first discriminant axis (according to the Land Use) as external drift.

On figure 5, 6 and 7 it is also shown the map of the initial normalised data.

2.5 Discussion

The analysis of the figures 5, 6 and 7 show that the introduction of an external variable into Factorail Kriging allows a better definition of the anomalous areas.

The analysis of the Cu maps (Figure 5) it is possible to identify three important regions: 1 and 3 located near the industrial areas (in the centre and to the SW of the Águeda municipality) and 2 located near the main population centres.

The regions 1 and 2 appear better defined and bigger on the C1 map, when using the Factorial Kriging with the land use information. The region 3 appears on that map (Figure 5 – Factorial Kriging with the land use information) with a completely different behaviour from the other maps, with high negative values. This is probably due to the fact that the samples from that area are classified (by the defined discriminant axis) as having forest or agricultural land use when in fact they are related to several industries.

The comparation of the Pb maps (Figure 6) it is possible to identify two important regions: 1 located near the industrial area (in the SW of the Águeda municipality) and 2 located near the Talhadas mine. These two regions appear in all maps but they have better expression on the C1 map of Factorial Kriging with the land use information, specially the region that corresponds to the Talhadas mine.

The analysis of the Zn maps (Figure 7) it is possible to identify one important region located near the urban and industrial areas (in the centre of the Águeda
Figure 5 – Maps of Cu initial standardized data and Factorial Kriging maps without external drift, with two types of external drifts: with lithological information and with land use information. The main urban settlements, the main heavy traffic road and the location of the Talhadas mine are shown.
Figure 6 – Maps of Pb initial standardized data and Factorial Kriging maps without external drift, with two types of external drifts: with lithological information and with land use information. The main urban settlements, the main heavy traffic road and the location of the Talhadas mine are shown.
Figure 7 – Maps of Zn initial standardized data and Factorial Kriging maps without external drift, with two types of external drifts: with lithological information and with land use information. The main urban settlements, the main heavy traffic road and the location of the Talhadas mine are shown.
municipality). This region appears in all maps but it is on the Factorial Kriging map with the land use information that it appears more extended and better defined.

The C0 maps of Factorial Kriging using the lithological information, for all variables, reflect the geological asymmetry present in this municipality.

3. Conclusions

The identified anomalies for Cu, Pb and Zn appear associated with the urban or industrial areas. The anomalies related to the industrial activities area mainly due to the electroplating and metal processing industries. The anomalies related to urban areas are explained by the presence of small, family enterprises.

The C0 maps of Factorial Kriging using the lithological information, for all variables, reflect the geological asymmetry present in this municipality.

The Factorial Kriging maps using the land use information as an external drift show the anomalous areas more extended and better defined probably due to the fact that the samples from those areas are classified (by the defined discriminant axis) as having forest or agricultural land use when in fact they are related to several industries.

The results show that the type of external drift affects the Factorial Kriging maps. The correlation between the type of external drift and the different variables is important for the enhancement of the anomalous areas.

4. Acknowledgements

The first author would like to thank the Fundação para a Ciência e a Tecnologia (FCT) the concession of a pos-PhD scholarship.
REFERENCES


