COMPARING SPLINES AND KRIGING USING SAS/GRAPH IN AN AGRONOMICAL APPLICATION.

CASALINO Francesco (*), CASTRIGNANO' Annamaria (**), ROMITO Lucia (*), ZUPA Vito (*).

(*) Centro Interdipartimentale di Servizio per l'Elaborazione ed il Calcolo Automatico - Universita' di Bari.
(**) Istituto Sperimentale Agronomico - Ministero Agricoltura e Foreste - Bari.

The work has been equally shared among the authors.

Introduction.

The soil changes from point to point, so it is important to introduce information about spatial variability in extrapolating soil properties into locations where they are unknown taking it from other ones where they are known. Classical statistical procedures assume that, about the mean, variability is random and contains no reference to the geographical position of the sample units. Several studies have shown that this "random" aspect of soil variability often contains a component that is spatially dependent (Castrignano et al. 1988; Castrignano and Lopez, 1990; Yost et al., 1982; McBranney et al., 1982): then the classical model is inadequate for the interpolation of spatially dependent variables.

Several techniques which incorporate sample locations have been used for the interpolation of soil properties. These include proximal weighing (Van Kuilenburg et al., 1982), moving averages (Webster, 1978), weighed moving averages using inverse distance and inverse distance squared functions (Van Kuilenburg et al., 1982), trend surface analysis (Whitten, 1975) and splines interpolation (Greville, 1969). These techniques are empirical, and although they may seem reasonable for many applications, they are theoretically unsatisfactory (Burgess and Webster, 1980). Most give biased estimates that are not optimal and many do not provide estimates of the interpolation error and those that do, do not attempt to minimize that error (Brugess and Webster, 1980).

Recent developments in statistical theory enable spatial dependence of soil properties to be directly considered in interpolation. This procedures, known collectively as "geostatistics", are based on "theory of the regionalized variables" (Matheron, 1971), which takes into account both the random and the structured characteristics of spatially distributed variables and provides an unbiased and optimal estimation. The interpolation based on spatial dependence of samples is called "Kriging" because it was first used by D. G. Krige (1951 and 1960) for the estimation of the gold content of ore bodies in the mining industry in South Africa. In particular, the aim of this work is to compare two methods: splines and kriging, at first from a theoretical point of view, then using a practical example.

Theory.

Splines.

When using spline interpolator, the most important goal is plotting contour lines as smooth as possible, so that the resulting map looks like what a draftsman would draw manually. This

SAS/GRAPH is a registered trademark of SAS Institute Inc., Cary, NC, USA.
propriety of the spline fittig can be obtained in both one and two dimensions by solving the
differential equation equivalent to a third order spline. This equation describes the displacement
of a stiff thin metal plate in one or two dimension forced to pass through the given data points.
The boundary conditions must be satisfied not only at the ends of the boundary, but also within
the region of interest. So, the solution is forced to take up the value of the observation at the
point of observation. The equation can be solved numerically and thus it gives a solution to
the problem of defining a set of values at the points of a regular grid. The smoothness properties
are a consequence of the method of deducing the difference equations, thus the smoothed
aspect of the resulting contour map is determined. The stiff thin metal strip (in one dimension)
or plate (in two dimensions) is bent by forces acting at the data points, so that the displacement
at these points is equal to the value of observation. If we denote (Briggs, 1974) the displacement
as $x$, the independent space variables as $x$ and $y$, the forces acting at $(x_n, y_n)$, for $n=1...N$, as
$f_n$ and the value of observations as $w_n$, then, in one dimension:

$$\frac{d^4u}{dx^4} = f_n \quad \text{if } x = x_n$$

$$= 0 \quad \text{otherwise}$$

and in two dimensions:

$$\frac{\partial^4u}{\partial x^4} + 2\frac{\partial^4u}{\partial x^2 \partial y^2} + \frac{\partial^4u}{\partial y^4} = f_n \quad \text{if } x = x_n \text{ and } y = y_n$$

$$= 0 \quad \text{otherwise.}$$ (2)

The solutions of the equations must satisfy the condition that $u(x_n) = w_n$ in one dimension
and that $u(x_n, y_n) = w_n$ in two dimensions. If, for simplicity, we consider the mono-
dimensional case, the displacement $u$, the slope $du/dx$ and the curvature $d^2u/dx^2$ must be
continuous at the point where the force is acting but $d^3u/dx^3$ is discontinuous at this point and
the value of the discontinuity is equal to the force acting (Love, 1926).
A solution is given by a third-order polynomial

$$u = a_0 + a_1x + a_2x^2 + a_3x^3$$ (3)

for each interval between the data points.
The coefficients of the polynomial are determined by the continuity conditions already
mentioned.
In two dimensions, the solution of equation (2) must be a bivariate polynomial in the
corresponding form with respect to the one (3) considered for the monodimensional case
(Hessing et al., 1972).

Kriging.

Kriging is performed in two steps (Matheron, 1973; Delfiner, 1975):

a) structural analysis:
in this step the spatially dependent component of the random variable z can be described by the semivariance \( \gamma (h) \), using the standard estimator:

\[
\gamma (h) = \left\{ \frac{1}{2n(h)} \right\} \sum_{i=1}^{n(h)} (z_i - z_{i+h})^2
\] (4)

where \( n(h) \) is the number of point couples separated by a distance \( h \), and \( z_{i+h} \) is the value of the variable at a point separated from the point \( i \) by a distance \( h \).

The interpretation of the semivariance values in eq. (4) is facilitated by fitting models to data using non-linear least-squares methods (Mulla, 1988).

The model that was the most appropriate for the data in our application was a gaussian model given by:

\[
\gamma(h) = C_0 + C_1 \left[ 1 - \exp \left( -\frac{|h|}{a^2} \right) \right]
\] (5)

where \( C_0 \) is a parameter known as the nugget, \( C_1 \) is a parameter representing the sill (asymptotic value of \( \gamma \) for \( h \to \infty \)) minus the nugget.

b) Kriging itself.

The value of the variable \( z \) at the nodes \( x \) of a regular grid is estimated by a linear combination of the values at \( n \) surrounding data points:

\[
z^*(x) = \sum_{i=1}^{n} \lambda_i z(x_i)
\] (6)

where \( z^*(x) \) is the estimated value at the point \( x \) and \( \lambda_i \) the weights.

The kriging estimator is the best linear unbiased estimator (B.L.U.E.), the weights are chosen so that the two following conditions are satisfied:
- the estimate \( z^*(x) \) of the true values \( z(x) \) is unbiased, i.e.:

\[
E \left[ z^*(x) - z(x) \right] = 0
\] (7)

The (7) simply means that kriging is an "exact interpolator": the estimated value is equal to the measured value at data points;
- the estimation variance \( \sigma_k^2 \) is minimized; i.e.

\[
\sigma_k^2 = \text{VAR} \left[ z^*(x) - z(x) \right] = \text{minimum}
\] (8)
These two conditions produce the following system of kriging equations:

\[
\begin{align*}
\sum_{j=1}^{n} \lambda_j \gamma(x_i, x_j) + \mu &= \gamma(x_i, x) \quad i = 1, \ldots, n \\
\sum_{j=1}^{n} \lambda_j &= 1
\end{align*}
\]  

where \( \gamma(x_i, x_j) \) and \( \gamma(x_i, x) \) are the semivariances between the observed location \( x_i \) and \( x_j \) and between the observed location \( x_i \) and the interpolated location \( x \) respectively; \( \mu \) is the Lagrangian multiplier associated to the minimization of \( \sigma_k^2 \). The solution of the \( n+1 \) equations of the kriging system, for each \( \lambda_i \) and \( \mu \), enables to estimate the kriged value of \( z^*(x) \) and the estimation variance:

\[
\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i, x) + \mu
\]

The set of \( n+1 \) simultaneous equations of the kriging system is most efficiently solved using matrix models (Burgess and Webster, 1980).

**Comparing kriging and splines.**

The objectives of the two methods are very different. The goal of kriging is to obtain a good average accuracy with a condition \( E[z^*(x) - z(x)] = 0 \) that guarantees against a systematical error. The second condition, i.e. the minimization of \( \text{VAR}[z^*(x) - z(x)] \) enables to obtain some estimates which are on the average, as close as possible to the actual values.

To apply kriging, one needs to carry out a stuctural analysis, that is an attempt of physical interpretation (modelling) of the phenomenon and allows to go beyond the simple description of the actual data, reflecting the effort of the researcher to understand the causes that produce that particular effect. Moreover, the objective of kriging interpolation is not to obtain a very fine map but an accurate one. The shape of \( z^*(x) \) is determined by the minimization criterion, whereas with the spline interpolation the shape of the interpolating function determines the minimization criterion. Besides, splines use the same analytical expression for the interpolator, whatever the structure of the variable, whereas kriging is based on modelling the experimental semivariograms that may take different forms, depending on the data and the sampling interval used. Splines may prove useful in particular when a quick visualization of the variable is needed, because they produce clear and fine maps which may prove inaccurate because the spline interpolator is not B.L.U.E. Then, splines must not be used to make calculations on reliable predictions. Finally, a useful feature of kriging is that an error term (estimation variance) is calculated for each estimated value, providing a measure of the reliability of the interpolation. Using splines, instead, interpolation is performed with no preliminary structural analysis; then, without knowing the structure of the variable, it is impossible to get an estimation variance.

**Agronomic application.**

The data set used contains 132 plot yield observations (t/ha) of durum wheat cropped during
the season 1984 - 85 in a field 101 m. x 47 m. in the experimental farm of the Istituto Sperimentale Agronomico, located in Foggia (Southern Italy). The data are referred to an agronomical test of continuous cropping of autumn-sown durum wheat that underwent four different tillage methods and three nitrogen doses. Sets of 10 random observations have been eliminated in turn from the original sample, to carry out 10 different tests.

Both interpolation methods, according to a regular grid with nodes m. 1.0 equally spaced, have been applied to the 10 resulting data sets - each one containing 122 observations - for a total of 4896 observations.

For splines the SAS/GRAPHG3GRID procedure for personal computers has been used, thanks to its capability to provide an exhaustive solution to the problem to be solved: it offers the possibility of choosing among three different methods of polynomial interpolation. In this case, the above-mentioned procedure has been used resorting to the option required to run the interpolation by splines.

With regard to kriging, the GEOPACK geostatistical software system has been used. This is a package of programs written by Dr Scott R. Yates, for conducting analyses of the spatial variability of one or more random functions. The system is menu driven and it produces a grid of estimates for the selected variable/s (*).

For kriging a previous structural analysis has been required according to which a gaussian model has been fitted to the yield semivariance.

To discuss the results, according to the tests carried out, the choice of that case, which has better allowed to point out the characteristics of each method, has been thought useful.

The parameters characteristic of the semivariogram model related to the considered case are reported in table 1.

Table 1: Output of structural analysis performed on one of the files deprived of 10 points.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>VALUE</th>
<th>STANDARD ERROR</th>
<th>t-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range (m)</td>
<td>48.55</td>
<td>7.124</td>
<td>6.82</td>
</tr>
<tr>
<td>Sill-Nugget [(t / ha)²]</td>
<td>.27</td>
<td>.027</td>
<td>9.91</td>
</tr>
<tr>
<td>Nugget [(t / ha)²]</td>
<td>.23</td>
<td>.017</td>
<td>13.72</td>
</tr>
</tbody>
</table>

The fitting process has been verified using the Jacknife technique (Efron and Gong, 1983), the results of which are reported in table 2.

Table 2: Results of Jacknife procedure.

<table>
<thead>
<tr>
<th>SUM OF SQUARES</th>
<th>REDUCED MEAN</th>
<th>REDUCED VARIANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>[(t / ha)²]</td>
<td>[t / ha]</td>
<td>[(t / ha)²]</td>
</tr>
<tr>
<td>5.149E-4</td>
<td>-3.354E-2</td>
<td>1.040</td>
</tr>
</tbody>
</table>
Then the SAS/GRAPH GCONTOUR procedure has been supplied with an input consisting of the data sets containing the values interpolated by the two methods, using the GXTDUMMY driver, provided by the Institute, to output the graphics coded in CGM format. This output has been used as input by a SAS programme (Casalino et al., 1990), able to code CGM format into DXF format, that is the standard recognized by most of CAD software. Such a choice has been suggested by the will of conciliating the use of SAS/GRAPH GCONTOUR procedure with software packages suitable to the thematic cartography for further employment in territorial statistics.

The yield contour lines map in fig. 1 has been obtained using interpolating spline, whereas the yield contour lines map in fig. 2 and the kriging variance contour lines map (fig. 3) have been obtained using kriging.

The splines map looks more regular, with smoother contour lines especially in extrapolation. Anyway, as mentioned before, the kriging interpolation offers two advantages compared to the splines one:
- it provides an estimate of the degree of reliability of the interpolation by the kriging variance map (fig. 3);
- the estimate provided by the kriging interpolation is more accurate than the one provided by splines.

To verify this, the following procedure has been used.

The 10 experimental data values taken away from the data set have been compared to the values interpolated using the two methods, thus the respective errors have been calculated.

To compare the two methods, an effective way is the evaluation of the estimating errors in the 10 values already mentioned, as shown in table 3.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Actual value</th>
<th>Spline value</th>
<th>Kriging value</th>
<th>Spline error</th>
<th>Kriging error</th>
<th>Kriging variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>1</td>
<td>3.58</td>
<td>4.19</td>
<td>3.68</td>
<td>-0.61</td>
<td>-0.10</td>
<td>0.2646</td>
</tr>
<tr>
<td>41</td>
<td>19</td>
<td>2.97</td>
<td>3.35</td>
<td>3.15</td>
<td>-0.38</td>
<td>-0.18</td>
<td>0.2627</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>2.51</td>
<td>3.07</td>
<td>3.19</td>
<td>-0.56</td>
<td>-0.68</td>
<td>0.2679</td>
</tr>
<tr>
<td>33</td>
<td>40</td>
<td>3.49</td>
<td>3.52</td>
<td>3.29</td>
<td>-0.03</td>
<td>+0.20</td>
<td>0.2626</td>
</tr>
<tr>
<td>33</td>
<td>50</td>
<td>3.54</td>
<td>3.67</td>
<td>3.23</td>
<td>-0.13</td>
<td>+0.31</td>
<td>0.2625</td>
</tr>
<tr>
<td>26</td>
<td>71</td>
<td>2.68</td>
<td>3.06</td>
<td>2.87</td>
<td>-0.38</td>
<td>-0.19</td>
<td>0.2624</td>
</tr>
<tr>
<td>26</td>
<td>82</td>
<td>2.89</td>
<td>3.13</td>
<td>2.93</td>
<td>-0.24</td>
<td>-0.04</td>
<td>0.2624</td>
</tr>
<tr>
<td>45</td>
<td>82</td>
<td>2.82</td>
<td>2.18</td>
<td>2.70</td>
<td>+0.64</td>
<td>+0.12</td>
<td>0.2676</td>
</tr>
<tr>
<td>17</td>
<td>92</td>
<td>2.77</td>
<td>3.10</td>
<td>2.95</td>
<td>-0.33</td>
<td>-0.18</td>
<td>0.2622</td>
</tr>
<tr>
<td>17</td>
<td>99</td>
<td>3.48</td>
<td>3.52</td>
<td>3.00</td>
<td>-0.04</td>
<td>+0.48</td>
<td>0.2635</td>
</tr>
</tbody>
</table>

The algebraic total of the errors is equal to -2.06 and -0.26 respectively for splines and kriging. As one can notice, for kriging such a result is closer to 0, that is a consequence of the fact that this method is unbiased, and this aspect preserves it from systematical errors.

Another way to compare the two methods is to calculate the sum of the squared errors, that is equal to 1.57 and 0.95 respectively for splines and kriging. This means that kriging is more accurate than splines. Besides, the greater errors (-0.61, -0.56, 0.64) have been obtained using splines, whereas using kriging just one great error (-0.68) has been found, corresponding also to a great error of splines.
Fig. 2: Kriging yield contour lines map
Fig. 3: Kriging variance contour lines map
Kriging and splines are two alternative methods depending on one’s objectives: the former is a good tool for predictions - the only one that should be used by researchers - whereas the latter quickly produces a clear map showing the main features of the studied variable.

Aknowledgements.

The authors wish to thank dr. Rizzo of the Istituto Sperimentale Agronomico - Bari for the permission of using data and Mr Giannotti of the Centro Interdipartimentale di Servizio per l’Elaborazione ed il Calcolo Automatico - Universita’ di Bari, for the desk-top publishing of the text.

(*) At the present the GEOPACK package is sold by GIBBS ASSOCIATES Energy and Mineral Specialists - P.O. Box 706 Boulder - Colorado 80306 - 0706.

Bibliography


