Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 91-115, 1998

Large-scale Versus Small-scale Variation Decomposition, Followed by Kriging Based on a Relative Variogram, in Presence of a Non-stationary Residual Variance

Laurent Raty
IGEAT, Université Libre de Bruxelles, CP 130/02, 50 Av FD Roosevelt, 1050 Brussels, Belgium
l_raty@hotmail.com

Marius Gilbert
Laboratoire de Biologie Animale and Cellulaire, Université Libre de Bruxelles, CP 160/12, 50 Av FD Roosevelt, 1050 Brussels, Belgium
mgilbert@ulb.ac.be

ABSTRACT **We propose here an interpolation method based on a decomposition of the data in large- and small-scale variation. This decomposition was performed using a two-way directional decomposition, similar to the decomposition used by Cressie in his median-polish kriging (1993), though we applied decomposition by means instead of medians. We considered the effects isolated by the decomposition as associated to intervals along the two directions, and not to a row or column co-ordinate. We replaced the "plating" used by Cressie to interpolate large-scale variation between grid rows and columns by the fit of a more continuous surface, with surface mean over each interval accounting for its associated mean effect.**

Residuals were predicted by kriging. However, as residual variance appeared to be strongly linked to large-scale-variation structure, we did not assume variance stationarity. Rather, we built a variance predictor based on this structure and modelled the data continuity through a relative variogram, relative to the sum of predicted variances at the tail and head of the lag vector. In the kriging equations, the variogram value used to characterize the variation between two locations was the relative variogram corresponding to the lag separating these two locations, rescaled by the sum of variance predictors at these two locations.

KEYWORDS: **trend, two-way decomposition, kriging, non-stationary variance, relative variogram.**

Contents

---

1. Foreword

When dealing with interpolation, the first question we can ask ourselves is: "What do we want to predict?" More or less outlying measures are often present among a data set. There are two "usual" ways to deal with them : (1) to identify them by some statistical procedure and to remove them from the data used to model the process which is supposed to be at the origin of the data; (2) to use so-called "robust" techniques, that will automatically downweight such data. The first technique is often considered too strong (except if the considered data clearly are errors in measurements or retranscriptions), as even outlying data may be thought to support information (see, e.g., Cressie 1993, p.144). The second method may seem more appropriate, but will anyway be the expression of our acceptance that some measures are wrong in some way, or do not respect the statistical model we chose to fit to them. Robust methods are *built to* downweight apparently aberrant data. Clearly, they will be unable to reproduce these data through the process of prediction. In this exercise, our purpose will be to predict the measures including, as much as possible, the more extreme ones. We have no theoretical reason to believe that some data are in error in any way, and we hope our models to be able to describe the more extreme as well as the more usual ones. We chose therefore deliberately not to use any robust method, hoping the complete data set may be adequately represented by a statistical model.

2. Decomposition of precipitation into large-scale and small-scale variation

2.1. Why did we choose to perform a decomposition?

Taking a look at the mapped data (Figure 1 – this map, as others in this paper, was generated by SURFER 5.01), even the inexperienced eye will detect an obvious

anisotropy in their behaviour. Any transect along the NW-SE direction will exhibit very strong variation in the precipitation, usually showing three minima surrounding two clear maxima. Variation along NE-SW transects is much less obvious.

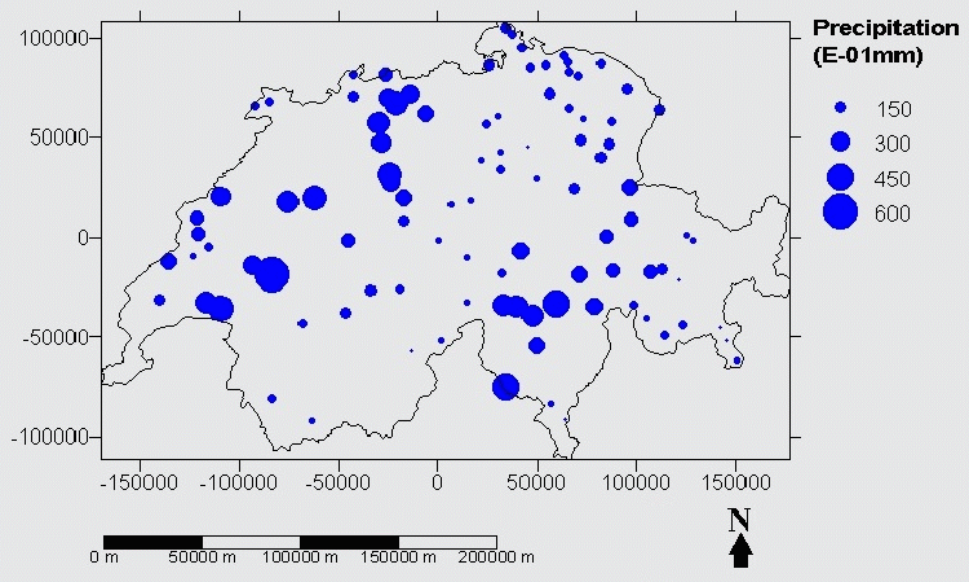

Figure 1. Map of available precipitation data (symbol radius is proportional to measured precipitation – this choice was made to enhance variation but may render some data points with very small precipitation value nearly invisible). Notice the fairly clear pseudo-periodic variation following NW-SE direction, with two obvious maxima surrounded with three minima. Variation along SW-NE direction is much less obvious.

This phenomenon has a strong influence on any attempt to build variograms on the basis of this data set : directional variograms do not behave in the same way along the various directions. In the NW-SE direction, data pseudo-periodicity leads to a semivariogram that increases largely above the data-set variance, reaches a maximum for lags corresponding to about a half-period of the variation, then decreases strongly at the largest lags. This behaviour is not present along the NE-SW direction, producing a more monotonous variogram.

According to Cressie (1993), decomposition of data into large-scale and small-scale variation is largely a choice of the statistician. Many problems can be modelled either by decomposing the process into sub-processes exhibiting more simple behaviours, or by making the covariance function used to describe the full process more complex. In the current application, adequate description of the process underlying the data without performing any decomposition would require to fit the directional variograms with different theoretical models, which causes transition problems between the two main directions. Furthermore, the pseudo-periodicity along the NW-SE direction occurs only on one and a half period, which is small to describe it in a statistical way. We believe that this periodicity at least would be best described as part of a trend, which led us to decompose the data into large-scale and small-scale variation. That is, according to Cressie's notations (1993), to assume :

$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s})$

where **s** is a vector denoting location, $Z(\mathbf{s})$ is the potential datum at this location, $E(Z(\mathbf{s})) = m(\mathbf{s})$ is a large-scale-variation deterministic structure, also referred to as the trend, and $d(\mathbf{s})$ is a zero-mean process, supposed to be stationary. There are two great types of methods that can be used to perform such a decomposition, that are both largely described by Cressie (1993). First, the large-scale variation can be assumed to follow a given mathematical expression : a spatially explicit mathematical model can then be fit to the data through some kind of regression. This leads to methods such as the well known "universal kriging". These methods always require quite strong assumptions as to the form of the model. Second, one can try to isolate the large-scale variation without making assumptions about its form. We chose to use a method rather similar to the one used in Cressie's median-polish kriging (Cressie 1993).

2.2. The theory beyond the proposed decomposition method

2.2.1. Two-way decompositions

Two-way decompositions of data into two main directional effects is not a new idea, one of the most fully documented application of it being "median-polish kriging", extensively described by Cressie (1993). Median polish was initially presented by Tukey (1977) as a fast and robust way to decompose gridded data into global, row and line effects. The principle is rather simple, and follows the idea of any two-way decomposition :

$Z(\mathbf{s}) = \text{"global effect"} + \text{"column effect"} + \text{"row effect"} + \delta(\mathbf{s})$

In practice, row and column medians are alternately extracted from the gridded data, subtracted to them to produce residuals, and accumulated into "row" and "column effects". Simultaneously, "column" and "row-effect" median are subtracted to these effects and accumulated into a "global effect". At any step of this operation, the sum for any data point of the "global effect", "row effect", "column effect" and of this datum residual equals the original datum. This algorithm is normally repeated until convergence. The method was later used by Cressie (1993) to isolate large-scale variation in gridded and non-gridded spatial data sets, and produce stationary residuals, suitable for kriging. The global method is referred to as "median-polish kriging". According to this author, any other averaging operation than median extraction will lead to its own version of the decomposition : median polish is just presented as an easy way to produce resistant estimates of the trend. As previously stated, we did not want to use particularly resistant methods, so we preferred to perform a decomposition by means, rather than by medians.

2.2.2. The problem of non-gridded data

Two-way analysis is straightforward and natural when dealing with gridded data, with data points already arranged in rows and columns. With non-gridded data, Cressie's recommendation (in the case of median polish) is to superpose a low-resolution grid to the data, to assign each datum to the nearest grid node, and to perform the median polish assuming the data were gridded. Linear interpolation and extrapolation, respectively within and outside the grid, are then used to achieve the definition of a large-scale variation surface, covering the whole spatial domain within which predictions are needed. At this stage, data are replaced at their original location, and residuals can be obtained by subtracting them the value of the median-polish surface at this location.

We can see two shortcomings in this method. First, "row" and "column effect" are, in reality, *not* associated to the grid row or column co-ordinates, but to *intervals* over which these rows or columns have an influence. A direct consequence is that, if a row or column corresponds to an extremum in "row" or "column effects", the large-scale variation associated to this row or column will always be averaged over the associated interval; linear interpolation will not take this effect into account and might yield locally non-stationary residuals. Second, statistical interpolation procedures such as kriging take advantage of data continuity and smoothness, and assume that the variable exhibits the same behaviour over the whole study area. If the data themselves are assumed to be smooth, subtracting them a surface that, being a juxtaposition of secant planes, exhibits slope discontinuities, will likely result in a loss of smoothness. This might yield residuals with larger variogram values, which would increase interpolation uncertainties. Furthermore, it is probable that residuals will behave differently away and close to the trend-surface discontinuities. These two problems will be most obvious if the trend exhibits strong extrema and if intervals are large (sparse data, leading inevitably to a very low-resolution grid).

2.2.3. Proposed solution

We will therefore try to propose an equivalent method that might avoid these problems. First, instead of superposing a grid to the sampled domain, we will divide it into intervals, following two main perpendicular directions, say x and y (in practice, these two methods are perfectly equivalent, but we prefer to avoid the concept of displacing data from their original location to a subjective grid-node). Two-way analysis may then be performed on the basis of x and y-classes and will produce exactly the same results as in Cressie's grid-based method. As we chose to make a decomposition by means, these results will be mean effects : estimates of the difference between the global mean and the mean over each particular x or y-class. Next step is to find a surface, defined over the whole domain, that will (1) ensure that the mean over each x- and y-class has the value isolated through two-way decomposition and (2) respect the assumed continuity of the large-scale variation. In the same time, we would like to preserve the advantages of the 2-way decomposition, that does not require any preliminary hypothesis about the global form of the large-scale variation. Let's concentrate on one x or y-interval. In Cressie's method, the trend effect over this interval is modelled by two straight-line segments, intersecting at the grid row or column co-ordinate. Instead, to avoid the discontinuity created by this intersection, we will try to describe this segment's trend effect with one and only one equation. The mean value of this equation over the interval has to equal the mean effect given by the decomposition.

Simultaneously, to ensure continuity between contiguous intervals, we will impose that both the values of this equation and of its first derivative at each interval extremity must equal those of the similar equations describing the trend over contiguous intervals. Now, if we defined n intervals along this direction, we are left with $3n-2$ constraints. There is no theoretical reason to select any particular equation to describe the large-scale variation over each interval, but we need at least as many parameters in the complete equation system as we have constraints. We chose to use polynomial expressions, mainly because they are easy to handle. In the aim to allow for some asymmetry and/or for the presence of an inflexion point within the trend in an interval,

we chose expressions of the 3rd degree. For n intervals, we will therefore introduce $4n$ parameters in the equation system: having more parameters than constraints, the system is not determined.

Constraints have to be added to reach only one solution. Our main idea here will be that we want, as much as possible, to avoid to "create" information (such information "creation" might happen, for example, if we described a constant mean with a periodic function: the mean of the model over each interval could indeed equal the mean of the process, but the periodicity is "created" information that does not exists in reality). To avoid this kind of problem, we propose to minimise the trend variation, over the whole sampled domain. An analytically easy way to achieve this result is to minimise the integral of the squared first derivative of the trend over the domain. This minimisation will act as additional constraints.

Rewrite the decomposition in terms of the data:

$$z_k(x) = \sum_{i=0}^{3} \beta_{ik}\, x^i$$

$$Z(\mathbf{s}_i) = z_a + z(x_{\mathbf{s}_i}) + z(y_{\mathbf{s}_i}) + R(\mathbf{s}_i)$$

The "global effect" z_a is estimated from the decomposition. We need a unique definition for $z(x)$ and $z(y)$ that would be valid over the whole sampled domain. The observed trend effect $<z_k>_{obs}$ over any x-interval k ranging from $x_{k\,min}$ to $x_{k\,max}$, will be modelled by a 3rd-degree polynomial expression :

Its first derivative is given by:

$$z'_k(x) = \sum_{i=0}^{3} i\, \beta_{ik}\, x^{i-1}$$

And the mean of this expression over the interval is given by:

$$< z_k > = \frac{1}{x_{k\max} - x_{k\min}} \sum_{i=0}^{3} \frac{1}{i+1} \beta_{ik} \left(x_{k\max}^{i+1} - x_{k\min}^{i+1}\right)$$

The constraints may therefore be written as follows :

$$\begin{cases} \dfrac{1}{x_{k\max} - x_{k\min}} \displaystyle\sum_{i=0}^{3} \frac{1}{i+1} \beta_{ik} \left(x_{k\max}^{i+1} - x_{k\min}^{i+1}\right) = < z_k >_{obs} & k = 1,\ldots,n \\ \displaystyle\sum_{i=0}^{3} \beta_{ik}\, x_{k\max}^{i} = \sum_{i=0}^{3} \beta_{i(k+1)}\, x_{(k+1)\min}^{i} & k = 1,\ldots,(n-1) \\ \displaystyle\sum_{i=0}^{3} i\, \beta_{ik}\, x_{k\max}^{i-1} = \sum_{i=0}^{3} i\, \beta_{i(k+1)}\, x_{(k+1)\max}^{i-1} & k = 1,\ldots,(n-1) \end{cases}$$

The first equation imposes equality between the mean of the model and the observed mean effect over each x-interval, the second one imposes continuity in the trend value between every pair of contiguous x-intervals, and the third one imposes continuity of the trend first derivative between these pairs of intervals.

The trend squared variation may be written:

$$\sum_{k=1}^{n}\int_{x_{i\min}}^{x_{i\max}}\left(\sum_{i=0}^{3} i\,\beta_{ik}\,x^{i-1}\right)^2 = \sum_{k=1}^{n}\sum_{i=0}^{3}\sum_{j=0}^{3}\frac{ij}{i+j-1}\,\beta_{ik}\,\beta_{jk}\,\left(x_{k\max}^{i+j-1} - x_{k\min}^{i+j-1}\right)$$

which is what we want to minimise. The constraints will be introduced in the equation system through the Lagrange parameter method. Therefore, instead of the last expression, we will minimise:

$$\begin{aligned}
&\sum_{k=1}^{n}\sum_{i=0}^{3}\sum_{j=0}^{3}\frac{ij}{i+j-1}\,\beta_{ik}\,\beta_{jk}\left(x_{k\max}^{i+j-1} - x_{k\min}^{i+j-1}\right) + \\
&\qquad \sum_{k=1}^{n}\lambda_k\left(\frac{1}{x_{k\max} - x_{k\min}}\sum_{i=0}^{3}\frac{1}{i+1}\,\beta_{ik}\left(x_{k\max}^{i+1} - x_{k\min}^{i+1}\right) - \langle z_k \rangle_{\partial\delta_3}\right) + \\
&\qquad\qquad \sum_{k=1}^{n-1}\mu_k\left(\sum_{i=0}^{3}\beta_{ik}\,x_{k\max}^{i} - \sum_{i=0}^{3}\beta_{i(k+1)}\,x_{(k+1)\min}^{i}\right) + \\
&\qquad\qquad\qquad \sum_{k=1}^{n-1}\nu_k\left(\sum_{i=0}^{3} i\,\beta_{ik}\,x_{k\max}^{i-1} - \sum_{i=0}^{3} i\,\beta_{i(k+1)}\,x_{(k+1)\min}^{i-1}\right)
\end{aligned}$$

where the l_k, m_k and n_k are respectively n, n-1 and n-1 Lagrange parameters. Minimisation may be achieved by differentiating this equation with respect to the b_{ik} and to the Lagrange parameters and setting the derivatives to zero. Differentiation with respect to the b_{ik} produces 4n equations, namely:

$$\begin{cases}
2\sum_{j=0}^{3}\frac{ij}{i+j-1}\,\beta_{i1}\left(x_{1\max}^{1+j-1} - x_{1\min}^{i+j-1}\right) + \frac{1}{i+1}\,\lambda_k\,\frac{\left(x_{1\max}^{i+1} - x_{1\min}^{i+1}\right)}{\left(x_{1\max} - x_{1\min}\right)} + \\
\qquad \mu_1\,x_{1\max}^{i} + \nu_1\,i\,x_{1\max}^{i-1} = 0 & i = 0,\ldots,3 \\
2\sum_{j=0}^{3}\frac{ij}{i+j-1}\,\beta_{ik}\left(x_{k\max}^{1+j-1} - x_{k\min}^{i+j-1}\right) + \frac{1}{i+1}\,\lambda_k\,\frac{\left(x_{k\max}^{i+1} - x_{k\min}^{i+1}\right)}{\left(x_{k\max} - x_{k\min}\right)} + \\
\qquad \mu_k\,x_{k\max}^{i} - \mu_{(k-1)}\,x_{k\min}^{i} + \nu_k\,i\,x_{k\max}^{i-1} - \nu_{(k-1)}\,i\,x_{k\min}^{i-1} = 0 & i = 0,\ldots,3;\ k = 2,\ldots,(n-1) \\
2\sum_{j=0}^{3}\frac{ij}{i+j-1}\,\beta_{in}\left(x_{k\max}^{1+j-1} - x_{k\min}^{i+j-1}\right) + \frac{1}{i+1}\,\lambda_k\,\frac{\left(x_{n\max}^{i+1} - x_{n\min}^{i+1}\right)}{\left(x_{n\max} - x_{n\min}\right)} - \\
\qquad \mu_{(n-1)}\,x_{n\min}^{i} - \nu_{(n-1)}\,i\,x_{n\min}^{i-1} = 0 & i = 0,\ldots,3
\end{cases}$$

while differentiation with respect to the Lagrange parameters simply reproduces the 3n-2 constraints. We are therefore left with a linear system of 7n-2 equations that can be

solved through one single matrix inversion. Determination of these parameters allows us to define uniquely $z(x)$. Similar calculations lead to the definition of $z(y)$, and an estimate of the large-scale variation is therefore given by :

$$\mathcal{M}(\mathbf{s}) = z_a + z(x_s) + z(y_s)$$

The residual $R(\mathbf{s}_i)$ may be obtained simply by subtracting the trend value to the data $Z(\mathbf{s}_i)$.

2.3. Application to the precipitation data

Our main aim when applying a large-scale *versus* small-scale variation decomposition was to isolate the pseudo-periodic behaviour of precipitation along the NW-SE direction. To ensure that this goal would be reached, we performed a 45-degrees rotation of the axes before to apply the decomposition itself. In the remaining part of this paper, the data-point locations will therefore be described in terms of northeasting and southeasting. The area was divided into 11 southeasting and 6 northeasting classes. The results of the two-way decomposition are summarised in Table 1. The above-described equation system was then applied, and produced the parameters required to describe the trend through 3rd-degree polynomial expressions. These parameters are summarised in Table 2.

Table 1 Two-way decomposition by means of the precipitation, following Southeast and Northeast directions. Lower and upper SE'ing and NE'ing indicate limits of intervals that divide the sampled domain following these directions

global contrib. (E-01mm)	SE'ing class # (k)	lower SE'ing (km)	upper SE'ing (km)	class contrib. (E-01mm)	NE'ing class # (k)	lower NE'ing (km)	upper NE'ing (km)	class contrib. (E-01mm)
176.71	1	-121	-85	34.00	1	-112.00	-77.50	5.00
	2	-85	-60	107.62	2	-77.50	-17.50	-5.00
	3	-60	-40	177.00	3	-17.50	22.50	18.50
	4	-40	-20	9.37	4	22.50	47.50	9.25
	5	-20	-5	-6.00	5	47.50	90.00	-52.75
	6	-5	10	-48.00	6	90.00	124.00	-15.00
	7	10	30	0.00				
	8	30	50	-11.25				
	9	50	75	130.75				
	10	75	115	-32.25				
	11	115	151	-44.25				

Figure 2a shows precipitation large-scale variation surface, modelled over the whole sampled domain. For comparison purposes, Figure 2b shows the trend that would have been obtained, should we have chosen to use Cressie's method of superposing a low-resolution grid, assigning the data point values to the nearest grid node, and performing linear interpolation. Differences between the two methods are mainly obvious at co-ordinates where there are strong discontinuities in Cressie's plating, and at sampled-domain extremities where Cressie's method relies on linear extrapolation based only on two values.

Table 2 Parameters used for modelling precipitation large-scale variation, following Southeast and Northeast directions, through 3rd-degree polynomial expressions. Lower and upper SE'ing and NE'ing indicate limits of intervals that divide the sampled domain following these directions. "Grid" SE'ing and NE'ing refer to the co-ordinates of grid rows and columns that would have led to the same division in Cressie's method (all data points in each interval are closer to this co-ordinate than to any other row or column co-ordinate).

SE'ing class #	**lower SE'ing**	**upper SE'ing**	**«grid» SE'ing**	**class contrib.**	b_{k0}	b_{k1}	b_{k2}	b_{k3}
(k)	**(km)**	**(km)**	**(km)**	**(E-01mm)**				
1	-121	-85	-100	34.00	67.5	-0.176	-.0221	-.000127
2	-85	-60	-70	107.62	828	10.3	-.0911	-.00115
3	-60	-40	-50	177.00	-1040	-39.1	-.183	.00239
4	-40	-20	-30	9.37	138	16.9	.413	.000678
5	-20	-5	-10	-6.00	-97.1	-5.51	-.0673	.00333
6	-5	10	0	-48.00	-86.8	-1.43	.348	.00422
7	10	30	20	.00	-152	11.1	-.181	-.00212
8	30	50	40	-11.25	449	-25.4	.246	.00189
9	50	75	60	130.75	-224	-22.0	.915	-.00749
10	75	115	90	-32.25	1810	-34.0	.152	-.134E-05
11	115	151	140	-44.25	-371	3.92	-.0127	.242E-07

NE'ing class #	**lower NE'ing**	**upper NE'ing**	**«grid» NE'ing**	**class contrib.**	b_{k0}	b_{k1}	b_{k2}	b_{k3}
(k)	**(km)**	**(km)**	**(km)**	**(E-01mm)**				
1	-112	-78	-110	5.00	-271	-7.49	-.0678	-.000207
2	-78	-18	-45	-5.00	29.1	1.42	.0122	-.0000123
3	-18	23	10	18.50	19.8	.366	-.0160	.0000551
4	23	48	35	9.25	38.6	-2.09	.0910	-.00150
5	48	90	60	-52.75	240	-6.42	.00565	.000340
6	90	124	120	-15.00	-1140	25.2	-.185	.000452

Residual small-scale variation was calculated by subtracting the large-scale variation to the data. Plotting these residuals against N-S, W-E, NW-SE and WS-NE directions did not show any remaining clear large-scale variation. According to Cressie (1993), plotting the residuals against $(x_s - \langle x \rangle)(y_s - \langle y \rangle)$ is a way to detect departures from additively of the two main effects (the slope of the plot would be indicative of the significance of a hypothetical quadratic term). Such a plot did not show anything significant either.

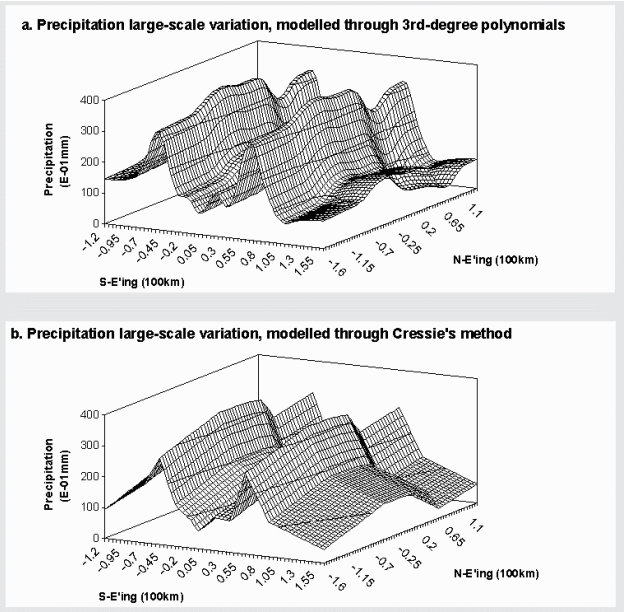

Figure 2 Precipitation large-scale variation isolated through two-way analysis by means (global effect + NE'ing class effect + SE'ing class effect), modelled (a) using 3rd-degree polynomials and (b) using Cressie's linear interpolation (for comparison).

2.4. Discussion

The method may appear complex but this is mainly due to the number of parameters that renders the equations fairly long. Its application is in fact fairly straightforward, and very easily programmed (the only "complex" issue is the inversion of large matrices, but any basic geostatistical library will anyway be furnished with procedures aimed for this purpose). The main software that we used was Microsoft Excel 97. Only inversion of large matrices was performed outside this program, using one of the matrix-inversion FORTRAN procedures furnished with GSLib (Deutsch and Journel, 1992). There are few hypotheses behind the model, except of course that large-scale variation can be decomposed into a global and two main directional effects (applicability of this hypothesis was verified through the last step of the decomposition analysis).

However, we had to take two subjective decisions during the process of developing the method, concerning respectively the analytical expression selected to describe the large-scale variation within each interval, and the quantity minimised to obtain a determined system. Third-degree polynomials were selected only to obtain easy-to-handle equations, allowing for the presence of extrema and/or an inflexion point within each interval. Anyway, we believe subjectivity is also present in Cressie's method, though maybe less obvious because the equations are apparently more simple. With trend effects associated with grid rows or columns, each one of them is viewed as a point located on one of the axes. Even if a straight line is, at first glance, the most simple way to join two points, the decision to choose this way to describe variation between these points is still completely subjective. Furthermore, if we look at the trend effects as being associated to intervals, rather than grid rows or columns, Cressie's large-scale-variation model within each interval is, in fact, composed by *two* secant straight-line segments, which is equivalent in complexity to a 3rd-degree polynomial expression (the number of

parameters is the same). Our decision to minimise the trend (squared) variation was largely dictated by the complexity of the large-scale variation and by the sparseness of the data points. Along the NW-SE direction at least, the trend is clearly far to be monotonous : even knowing its mean values within several contiguous intervals, one can hardly make any assumption about what could happen in the next interval. In such conditions, we preferred to model the trend behaviour by a smaller variation, rather than to make hypotheses about its continuity from the previous intervals. Influence of this choice is most obvious at the extremities of the analysed spatial domain, where the model behaviour is less influenced by mean effects in contiguous intervals, and therefore more influenced by the minimisation : our decision will invariably lead to a small trend slope at these extremities (which clearly differs from the linear extrapolation performed on the basis of the two outer rows or columns in Cressie's method). There is no way to assess the value of this decision on a general basis and without knowledge of the large-scale variation behaviour outside the sampled domain. At least one of its merits is that, when working with a variable that is bounded in some way (like precipitation, that can not be negative), the risk to see the large-scale variation fall outside the theoretical range of values for the analysed variable is smaller. Under different circumstances, for example in presence of a monotonous trend or with a larger amount of data, it might make more sense to minimise the variation of the trend first derivative rather than the variation of the trend itself, which would produce a modelled surface closer to the surface defined by Cressie's method.

One final word may be written about extrapolation. Would it be necessary to extrapolate outside the limits of the sampled domain, our recommendation might be to continue the trend linearly, using the trend slope at the domain limit. However, it must be understood that the method is intended for interpolation, *not* for extrapolation.

In the decomposition model, the trend is supposed to be deterministic; only the residual variation is analysed on a real statistical basis. In such models, it is very important for the trend values at any location to be estimated through the same process : even if a statistical procedure (kriging for example) might produce supposedly correct estimates of the small-scale variation outside the sampled domain, this small-scale variation would *not* be an estimate of residuals from an *extrapolated* large-scale variation. This remark, of course, is also valid for Cressie's method, though not emphasized by this author.

3. Geostatistical analysis of precipitation residual small-scale variation

3.1. Structural analysis

3.1.1. Residual variance

Once the trend has been removed, the classical method is to treat residuals $R(\mathbf{s}_i)$ as a new, independent and stationary data set. However, here, we ran rapidly into problems. Figure 3 shows residuals plotted against trend values. We can see that mean residual precipitation is slightly positive and, indeed, does not depend on trend values. Residuals appear even to be remarkably symmetrically distributed for any trend value. However, and this is probably a common situation with data bounded to zero, the residual variance is dependent on the trend value. As a result, the cloud span clearly increases to the right of the plot, even though data density decreases.

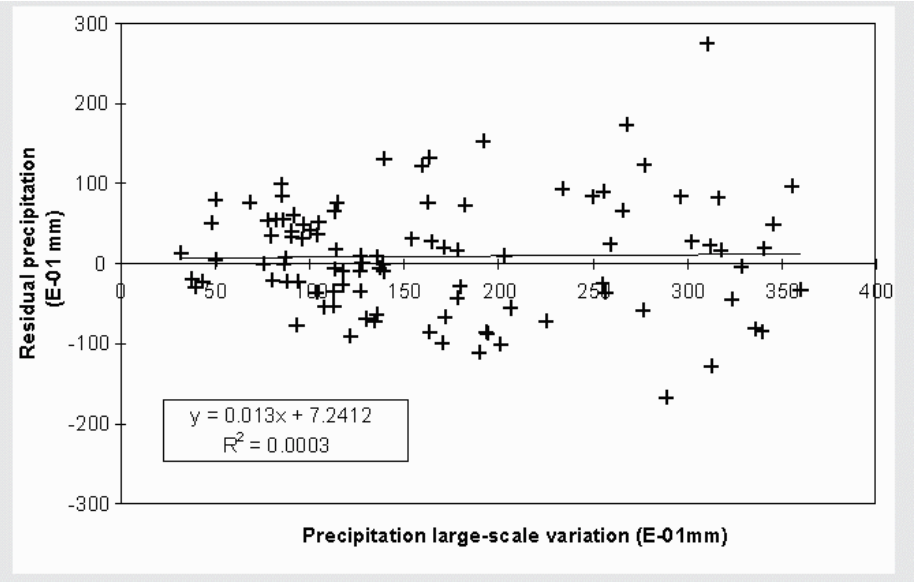

Figure 3 Residual precipitation plotted against precipitation large-scale variation isolated through two-way analysis by means. As shown by the regression line, the mean residual precipitation are not dependent on the large-scale variation. Residuals appear even to be fairly symmetrically distributed around their mean at all values of the large-scale variation. However, residual variance increases clearly with large-scale variation.

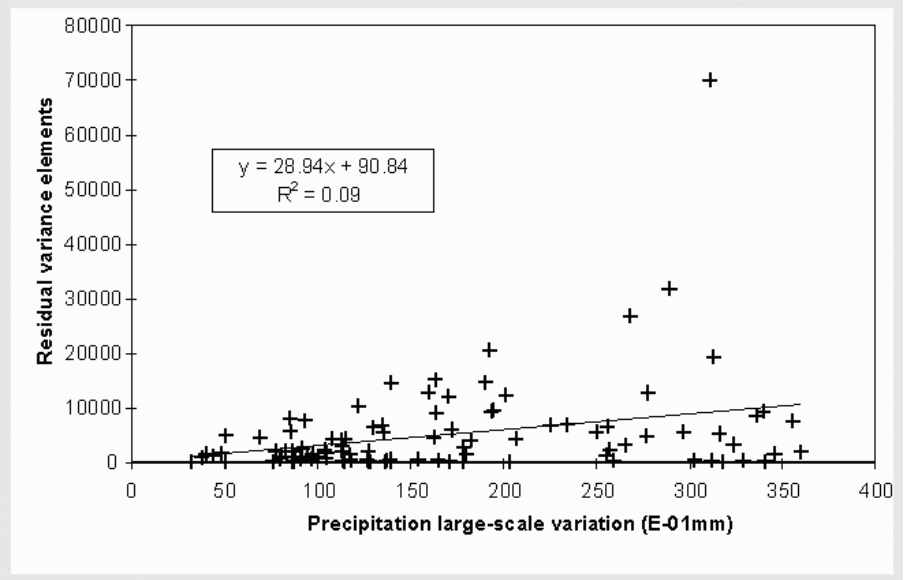

Figure 4 Precipitation residual variance elements plotted against precipitation large-scale variation isolated through two-way analysis by means. Residual variance elements tend to be proportional to the trend value (if the regression line is forced through the origin, equation becomes y = 29.36 x without significant loss in R^2).

On the basis of (1) the absence of any detectable residual trend in any direction and (2) the apparent independence between residual mean and large-scale variation, we will assume that residuals have a perfectly stationary mean, and that only their variance varies. Under this assumption,

$$s^2(\mathbf{s}_i) = (R(\mathbf{s}_i) - \overline{R})^2$$

is an element of the local residual variance at the location $\boldsymbol{s}$, and we can recombine these elements to calculate residual variances or plot them against any explanatory factor, to build an expression that could be used to predict the residual variance at any location. Dependence between residual variance and precipitation trend is clearly confirmed by the plot shown in Figure 4. Furthermore, the relation is very close to proportionality. One might, therefore, predict residual variance using the expression ($m(\mathbf{s})$ denotes the trend value in E-01mm) :

$$\hat{\sigma}^2(\mathbf{s}) = 29.36\, m(\mathbf{s})$$

One will notice, however, that proportionality is quite tempered at the highest values of the trend. In an attempt to find another explanatory factor, we calculated the trend gradient for each data point. Residual variance elements appeared to be positively correlated to this parameter as well, though this relation was much less clear ($R^2 = 0.07$). This result was to be expected for two reasons : first, we might expect the residuals to be more variable by themselves at locations where the precipitation are globally more variable; second, we can expect the trend fit to be less correct where its gradient is higher (a slight modification in the trend parameters would lead to a significant variation in the trend values that were subtracted to the data). Finding a global expression for residual variance is not straightforward, however, because (1) the two explanatory variables are significantly (though not strongly) positively correlated, and their respective effects are therefore "clouded" by each other, and (2) variation of variance elements increases with their mean, making classical linear regression procedures not really correct (more weight is systematically given to high-valued data). As proportionality to the trend value was the most clear influence, it was assumed that this was the main effect, and that influence of the trend gradient occurred only through a moderation or increase of the proportionality factor. Therefore, we plotted the ratio of variance element to the trend value, calculated for each point, against the trend gradient, and performed an ordinary least square regression. Thanks to the proportionality of the two terms of the ratio, this variable exhibits a relatively homoskedastic behaviour. This produced the following expression :

$$\hat{\sigma}^2 = m(s)\,(20.484\, grad(m(s)) + 19.614)$$

where $m(\mathbf{s})$ denotes the trend value (in E-01mm) and $grad(m(\mathbf{s}))$ its gradient (in mm/km). This expression explains 12% of the residual variance elements, which is already an achievement considering that fitting was performed on elements, that are highly variable quantities (when replacing $m(\mathbf{s})$ and $grad(m(\mathbf{s}))$ by their respective means, the same expression will predict respectively 97% of the variances calculated on classes of 20 data points with increasing trend values, and 83% of the variances calculated on classes of 20 data points with increasing trend gradient).

3.1.2. Stationarity assumptions

In the previous paragraph, we already made the assumption that residual mean was stationary. Under this assumption, we saw that precipitation residual variance, on the

contrary, is obviously *not* stationary, but is strongly linked to the precipitation large-scale-variation structure. Therefore, classical stationary assumptions do not really make any sense. Indeed, define the residual variogram as:

$$2\gamma(\mathbf{h}) = \mathrm{var}(R(\mathbf{s}) - R(\mathbf{s}+\mathbf{h})) = \mathrm{var}(R(\mathbf{s})) + \mathrm{var}(R(\mathbf{s}+\mathbf{h})) - 2\,\mathrm{cov}(R(\mathbf{s}), R(\mathbf{s}+\mathbf{h}))$$

Assuming $R(\mathbf{s})$ still behave approximately like a regionalized variable, when $\mathbf{h}$ becomes large, the covariance term tends to zero and 2γ tends to the sum of the residual variances at locations $\mathbf{s}$ and $(\mathbf{s}+\mathbf{h})$. If these residual variances are location-dependent, it is evident that 2γ can not depend only on $\mathbf{h}$. At least, its sill will be a function of locations of both the head and tail of the *vector* $\boldsymbol{h}$. It has been shown (see Cressie 1993, pp. 142-143) that kriging does not require variogram or covariance function to be stationary. In fact, as long as 2γestimates the statistical distance between the values taken by the variable at two locations $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$(that is, as long as it is an estimate of the variance associated to the difference between these values), kriging equation system remains valid and will still produce an optimal unbiased linear predictor. However, except perhaps in some cases where a definite physical model might suggest a particular covariance function, some stationarity assumptions are always necessary to allow the variogram or covariance function to be estimated. To solve the current problem, we propose to define the following (pairwise) relative variogram :

$$2\gamma_R(\mathbf{h}) = \frac{\mathrm{var}(R(\mathbf{s}) - R(\mathbf{s}+\mathbf{h}))}{\mathrm{var}(R(\mathbf{s})) + \mathrm{var}(R(\mathbf{s}+\mathbf{h}))}$$

with a theoretical sill of 1. The shape of $2\gamma_R(\mathbf{h})$ will be assumed to be the same everywhere in the sampled domain (only the variogram scale is supposed to be location-dependent). Similar assumptions, although not used to describe a variable continuity within a given spatial domain, have been made to build single relative variograms from data issued from separate sub-domains, that were assumed to show different variation scales (Isaaks and Srivastava 1989,Cressie1993). The residual relative variogram will be assumed to depend only on $\boldsymbol{h}$, while residual variance $\mathrm{var}(R(\mathbf{s}))$ will be supposed to be non-stationary, but dependent only on the large-scale variation structure (and, therefore, suitably predictable from this structure).

These assumptions are in fact fairly similar to the more classical assumption of second-order stationarity, except that the stationarity is here expressed in terms of relative covariance instead of covariance (that is, for any given lag $\boldsymbol{h}$, covariance is supposed to reduce the global variability of the difference between two data points separated by this lag by a given percentage – we propose to refer to this assumption as "relative stationarity").

3.1.3. Variography

The most natural estimator of the above-defined relative variogram is an analogous of David's pairwise relative variogram (1977), each squared difference in the classical Matheron's variogram estimator (1962) being rescaled by the sum of the predicted local variances at the head and tail of the vector $\mathbf{h}$:

$$2\hat{\gamma}_R(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\frac{(R(\mathbf{s}_i) - R(\mathbf{s}_j))^2}{\hat{\sigma}^2(\mathbf{s}_i) + \hat{\sigma}^2(\mathbf{s}_j)}$$

where the sum is over all data pairs separated by vector **h**. Similarly, under the assumptions defined in the previous paragraph, the continuity might also be described in terms of a relative autocovariance function :

$$\hat{C}_R(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|}\sum_{N(\mathbf{h})}\frac{(R(\mathbf{s}_i) - \bar{R})(R(\mathbf{s}_j) - \bar{R})}{\hat{\sigma}^2(\mathbf{s}_i) + \hat{\sigma}^2(\mathbf{s}_j)}$$

where the sum is again over all data pairs separated by vector ***h***. Still under these assumptions, definitions of these two parameters lead to the relation :

$2\gamma_R(\mathbf{h}) = 1 - 2C_R(\mathbf{h})$

which allows us to define $1 - 2\ C_R(\mathbf{h})$ as a "variogram form" of the covariance function.

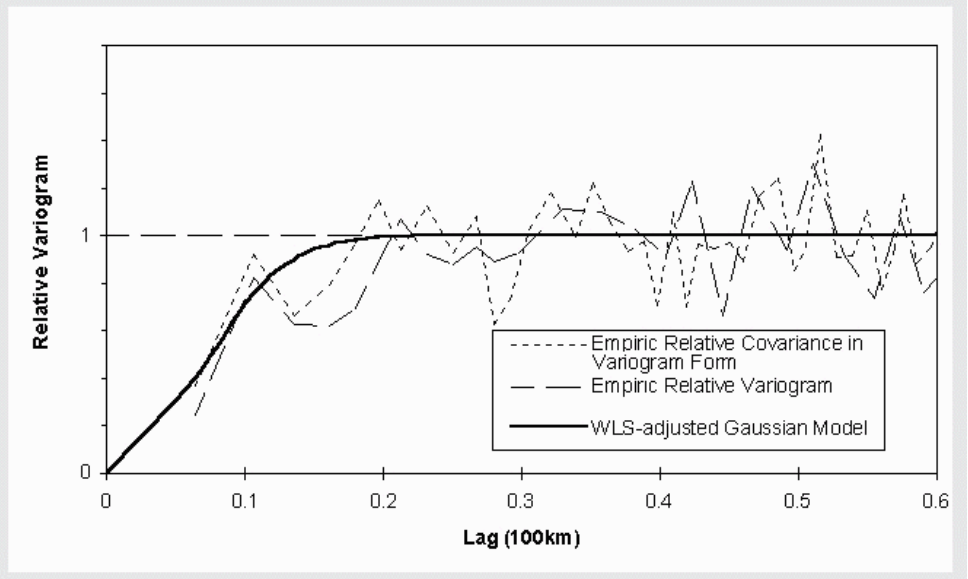

Figure 5 Empirical relative variogram ($2\gamma_R(\mathbf{h})$), relative covariance in variogram form ($1 - 2\ C_R(\mathbf{h})$) for the residuals, and Gaussian model adjusted to the covariance function in variogram form through Cressie's weighted least square method. Notice that both variogram estimators give similar results and that relative variogram indeed tends to 1.

Empirical relative variogram and relative autocovariance were computed using these two formulae; results are shown in Figure 5. One should notice that both parameters give very similar results and that relative variogram indeed tends to a sill of 1 (which might not have been the case, if the stationarity in mean had not been satisfied or if the residual variance predictor had acted poorly). A Gaussian variogram model with a sill of 1 was adjusted to the relative covariance in variogram form, using Cressie's weighted-least-squares method (1993). Best fit was obtained for a nugget effect of zero and a range of 15.5 km.

3.2. Estimation of residual small-scale variation

The residual-precipitation small-scale variation was estimated for each unknown data point through kriging. In this exercise, we used a classical kriging system, except that

variogram values characterising variation between each location pair were obtained from the relative variogram, rescaled by the sum of the residual variance predictors at these two locations. We did not reduce the kriging neighbourhood, and used all 100 available data for each point estimation. This produced estimates for residual precipitation small-scale variations, as well as kriging variances associated to these estimates. The kriging itself was performed using a modified version of one of the FORTRAN procedures furnished by GSLib (Deutsch and Journel, 1992).

3.3. Discussion

The use of relative variograms has been discussed to some extent by Isaaks and Srivastava (1989). Amongst other things, these authors proposed a case study of local-uncertainty assessment using a relative variogram, produced by rescaling the absolute variogram model in such a way that the sill equalled 1 (pp. 519 *et seq.*). They claim that error variances produced by kriging using this variogram are relative to the local variance : actual error variances associated to their estimates are then obtained by rescaling this variance with a local-variance estimate.

In practice, this procedure is perfectly equivalent to a direct rescaling of the kriging error produced by using an absolute variogram, simply by multiplying it with the ratio of the local variance to the absolute-semivariogram sill, without even introducing the relative variogram concept. We do admit that, in case of non-stationary variances, this will provide apparently more reasonable uncertainty assessment (actual errors are more likely to be correlated to the rescaled variance). Unfortunately, we can not agree with the theory underlying this practice.

Kriging produces best unbiased linear predictions thanks to a minimisation of the expectation of the prediction-error variance, which is performed here under the assumption that variance actually *is* stationary. To remove this assumption on the next step of the procedure is to accept that what was minimised in the first step actually was *not* representative of the error variance. Through the variogram, kriging equations take account of two sets of statistical distances, respectively between each data point and the estimation point, and between each data-point pair. If the first set of statistical distances may be thought to depend on the local variance at the prediction location, they can also be said to depend on the local variance at the data points. Statistical distances separating each data-point pair are, for them, not at all related to the local variance at the prediction location. Unless kriging is performed with a very restricted kriging neighbourhood within which the local variance may be assumed stationary, (1) the produced predictions will *not* be best linear predictions and (2) the produced error variance will *not* even be a correct estimate of the error variance associated to this prediction.

The method we proposed here is quite different from the above-described uses of relative variograms, in the way that our relative variogram is rescaled, by local variances *at the two* concerned points, *prior to* perform kriging. We strongly believe that this, only, leads to a set of equations taking into account the statistical distances between each data-point pair. There is a price to pay, however : under the assumptions we adopted, it is very difficult to *ensure* positive-definiteness of the model : rescaling the variogram theoretically destroy its positive-definiteness. This problem is difficult to avoid as the trend and, therefore, the residual variance, are largely experimental results. There would be two situations in which the model would be really positive-definite : (1) if residual

variance was stationnary and (2) if residual reduced variogram was a pure nugget effect. Neither of these situations seems to occur here. However, we do not believe that assuming that, practically, the model will behave as if it was positive definite is to take a great risk : (1) residual variance does vary, but smoothly and it is bounded to a limited interval and (2) residual auto-correlation is limited to a short range, as compared to the mean distance between data points. Any point farther than 15 km from a data point will behave as if the model was a pure nugget. Furthermore, the Gaussian model selected for variography is very smooth, which should ensure an efficient screen effect.

4. Estimation of total precipitation

Summing residual estimates to the trend for each location yielded precipitation estimates. Associated 95%-confidence intervals were built under a local normality assumption (estimated value ± 1.96 kriging standard deviation). Contour maps of the estimates and their associated standard deviation are shown respectively on Figure 6 and 7. Figure 6 also shows the locations of the 10 smallest and 10 largest estimated values.

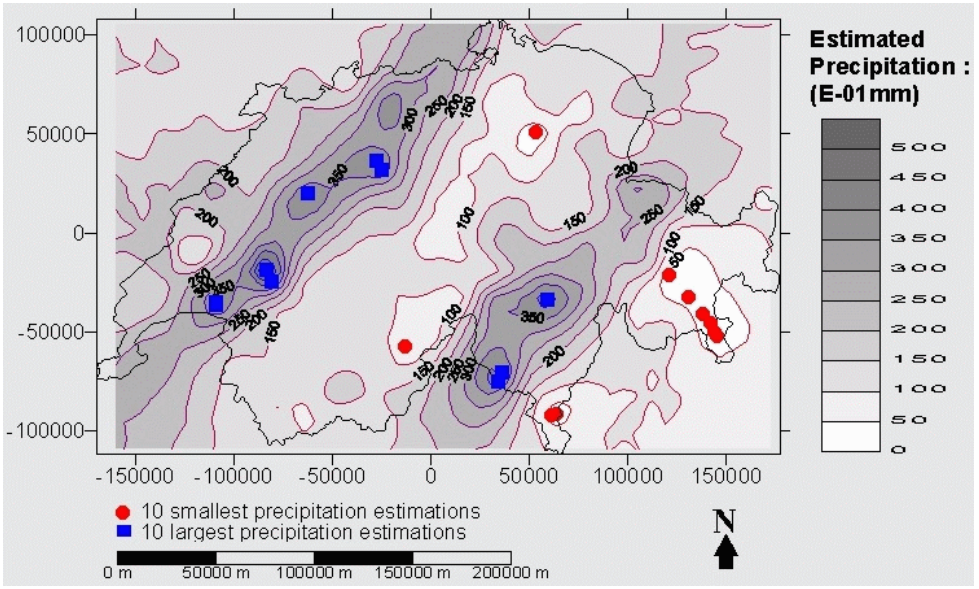

Figure 6 Contour map of estimated precipitation. The 10 largest and 10 smallest estimated values are also located on the map.

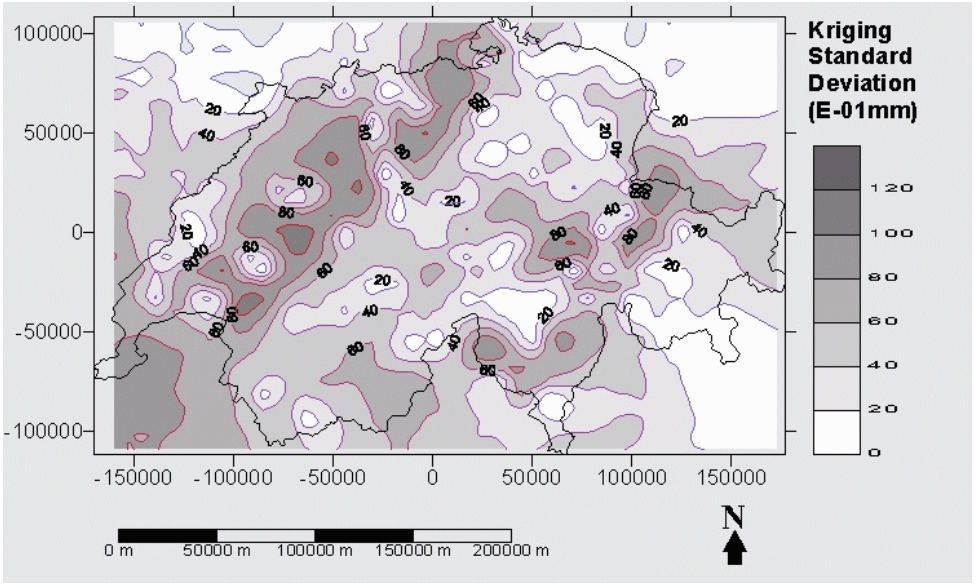

Figure 7 Contour map of kriging standard deviation associated to the estimation.

5. Method performances

5.1. Global comparison of estimated precipitation values versus actual precipitation values

Figure 8 shows true precipitation values plotted against our estimates, and Figure 9 gives a comparison of distribution histograms of true and estimated values. Table 3 summarises distribution parameters of true and estimated values, and gives several measures that can be used to assess performances of the method. Estimated values correlated very highly significantly with actual values (R^2 = 0.63), but there was a slight negative global bias (about 1mm, approximately 5% of the mean).

Table 3 Distribution parameters for actual and estimated precipitation, and prediction performances (bias : mean error; MAE : Mean Absolute Error; MRAE : Mean Relative Absolute Error; MSE : Mean Squared Error; RMSE : Root Mean Squared Error; R^2 : determination coefficient). Estimated values have narrower distribution than actual ones. Precipitation mean and median were both underestimated by about 1mm.

	True	Estimated	Performance Measures	
Minimum (E-01mm)	0.00	16.16	bias (E-01mm)	-10.2
Maximum (E-01mm)	517.00	435.38	MAE (E-01mm)	50.8
Mean (E-01mm)	185.36	175.21	MRAE (%)	34.3
Median (E-01mm)	162.00	153.62	MSE (E-02mm^2)	4711.6
Variance (E-02mm^2)	12358.18	7278.51	RMSE (E-01mm)	68.6
			R^2	0.63

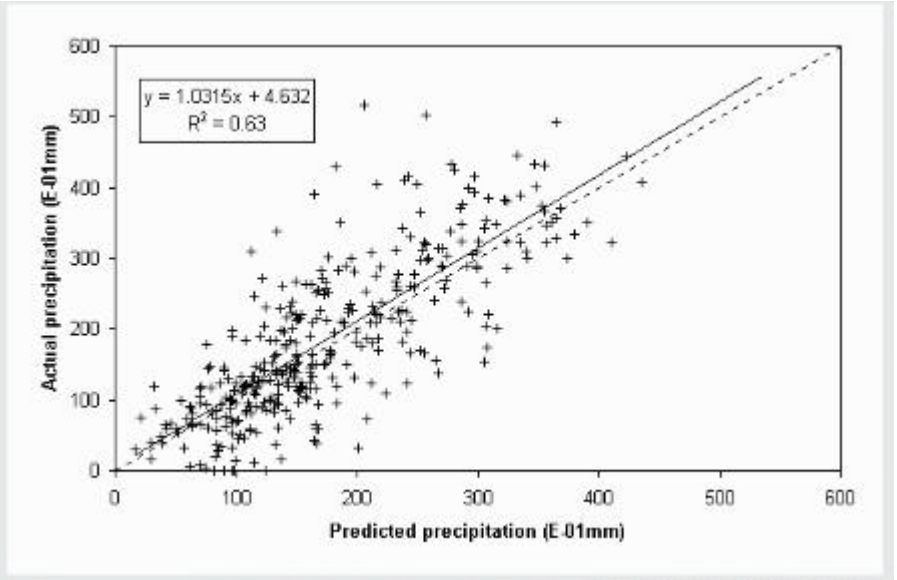

Figure 8 Actual precipitation values plotted against estimated values. The lines through the plot represent respectively y = x and a linear regression. The correlation is highly significant, the slope do not differ significantly from 1, nor does the intercept differ significantly from 0.

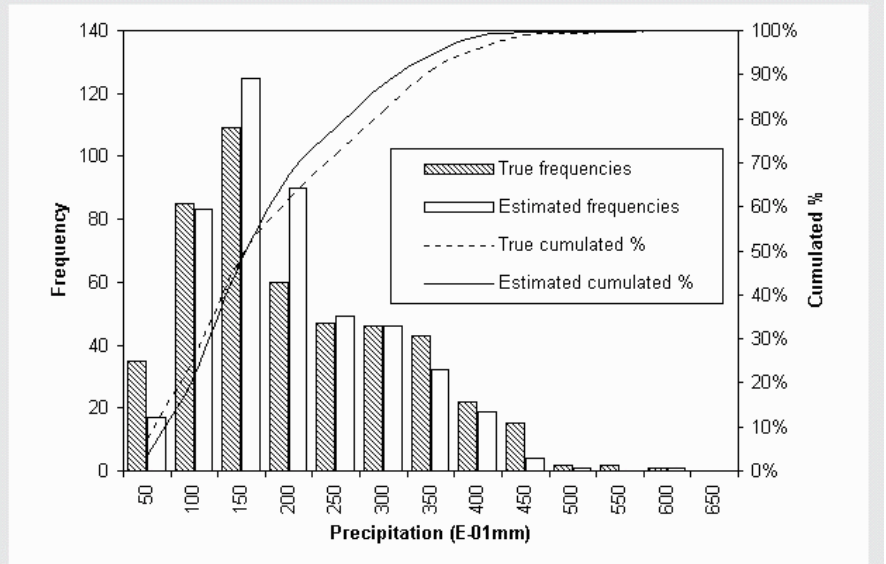

Figure 9 Distribution comparison of true precipitation values versus predicted values. Distribution shapes are quite similar, but one can easily notice the "smoothing" effect of the estimation procedure, yielding frequencies slightly higher in central classes and smaller in lower and upper classes than in reality.

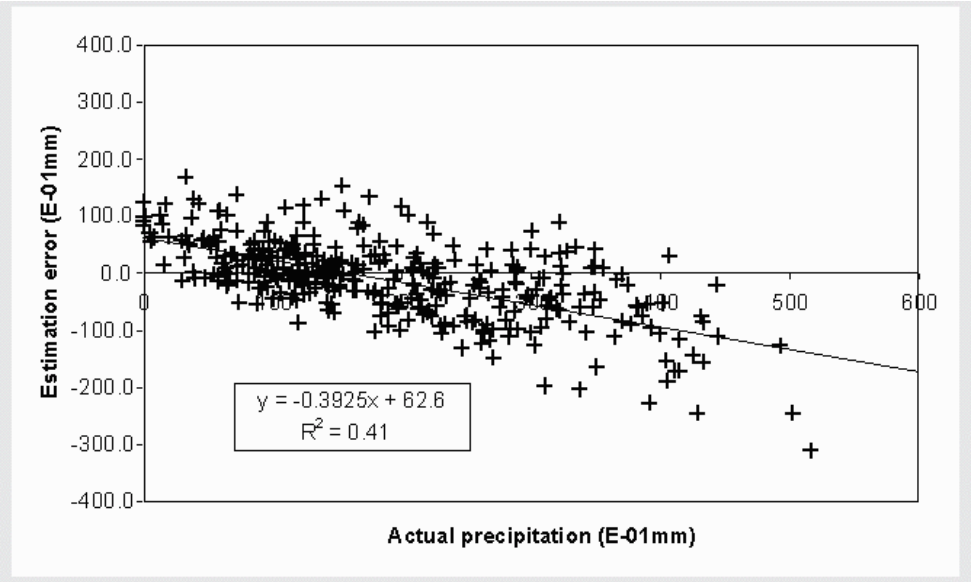

Figure 10 Conditional bias caused by the "smoothing" effect of kriging. Errors are obviously negatively correlated to the true precipitation values.

One inconvenience of the method was conditional biasedness : it is a fairly known phenomenon that kriging acts like a "smoother", yielding estimates exhibiting a more narrow distribution than actual data values. In fact, this problem is shared by most interpolation methods. Extreme predictions are usually obtained only at location for which high weight is given to one or more extreme-valued data, and high weights normally are the expression of a fairly good level of certainty in the prediction. It is therefore relatively rare for extreme predictions to be highly erroneous. On the other hand, less precise predictions are usually closer to the mean (more similar weight is given to all data), but some of these predictions may still be representative of true extreme values. Therefore, when taking into account only prediction values, extreme data are less often predicted than they do occur in reality. This effect can easily be seen in Table 3, estimated precipitation values showing higher minimum value, lower maximum value and lower variance than the actual precipitation data. The main consequence is that actual largest values are often underestimated, while actual smallest values are more likely to be overestimated. A regression of precipitation errors against actual precipitation values yielded indeed a very highly significant negative slope of -0.39 (R^2 = 0.41; see Figure 10).

5.2. Comparison of assessed uncertainty versus actual errors

Actual error magnitude is clearly linked to kriging standard deviation, as shown by the scatterplot on Figure 11. Furthermore, relation is close to proportionality, with a proportionality factor not too far from 1 (a much stronger relationship between these two parameters is not to be expected (Isaaks and Srivastava, 1989)). This is already a very

good point, and shows that kriging standard deviation may be used as an indicator of local uncertainty associated to our method.

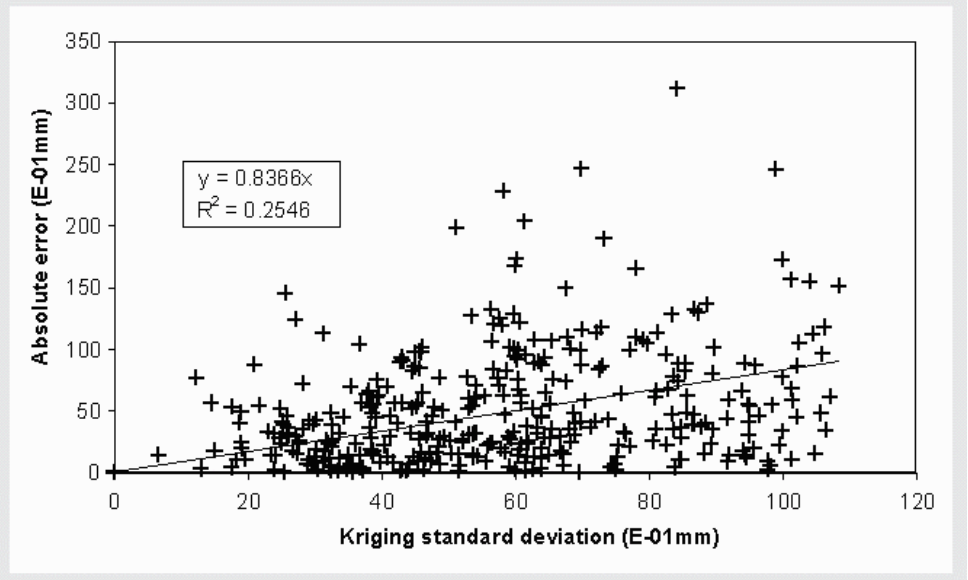

Figure 11 Scatterplot of actual absolute errors versus kriging standard deviation. The magnitude of actual errors clearly increases with increasing kriging standard deviation.

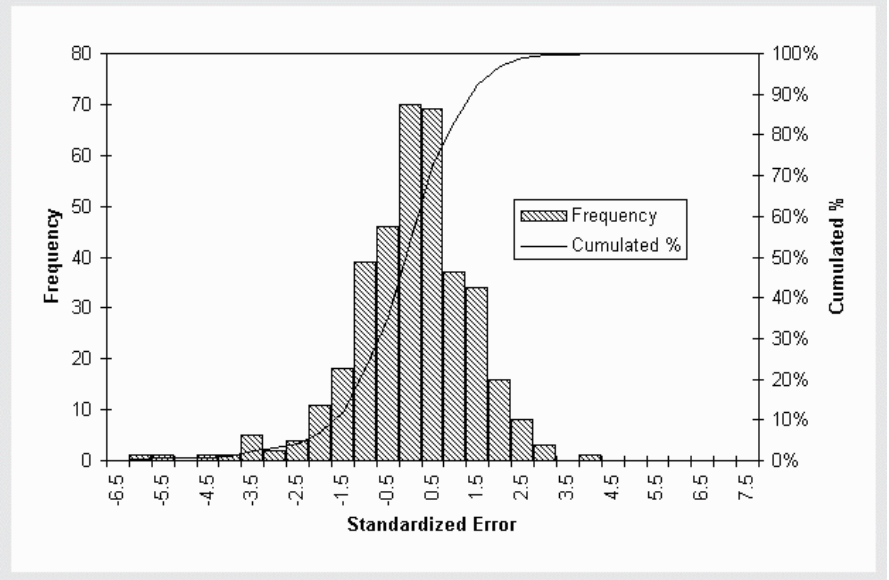

Figure 12 Distribution histogram of standardised errors ((estimated value– true value) / kriging standard deviation). Notice that histogram is centred on a value that is close to zero (bias is small), and is relatively symmetrical, except for a few strongly negative values.

A histogram of standardised errors (divided by the kriging standard deviation) is shown on Figure 12, and Figure 13 shows a map of raw errors, using proportional symbols. The histogram is centred on a slightly negative value (mean : -0.19, median : -0.11), and roughly symmetrical. However, a handful of data are quite strongly underestimated (these data show on the left of the histogram as a small secondary mode, around -3.5). As a consequence, the standard deviation of standardised errors is 1.28, thus quite higher than the ideal value of 1. The proportion of true data falling outside the 95%-confidence intervals built under a normality hypothesis was 10.6% : respectively, 3.5% fall lower than the lower confidence limits (overestimation), and a well higher 7.1% fall higher than the upper limit (underestimation).

In fact, such results are partly to be expected. Remember large-scale variation was isolated from the data set. Should some "anomaly" occur, leading to a higher mean in an area without data to show it, there would be no way to isolate this variation : actual residuals in this area would not be stationary in mean (there would be an (undetectable) "pocket" of non-stationarity), and precipitation would likely be underestimated. But this would not be the only consequence... Local residual variance, as a function of local trend value, would be underestimated as well, producing an associated confidence interval *too narrow*. The opposite situation, with a lower local mean than expected, would lead to overestimate the precipitation. However, this would be accompanied by an overestimation of local residual variance, and confidence interval would therefore likely be *too large*, making error appear less significant. Structurally, our model leads therefore to higher risk to "significantly" underestimate data than to "significantly" overestimate them.

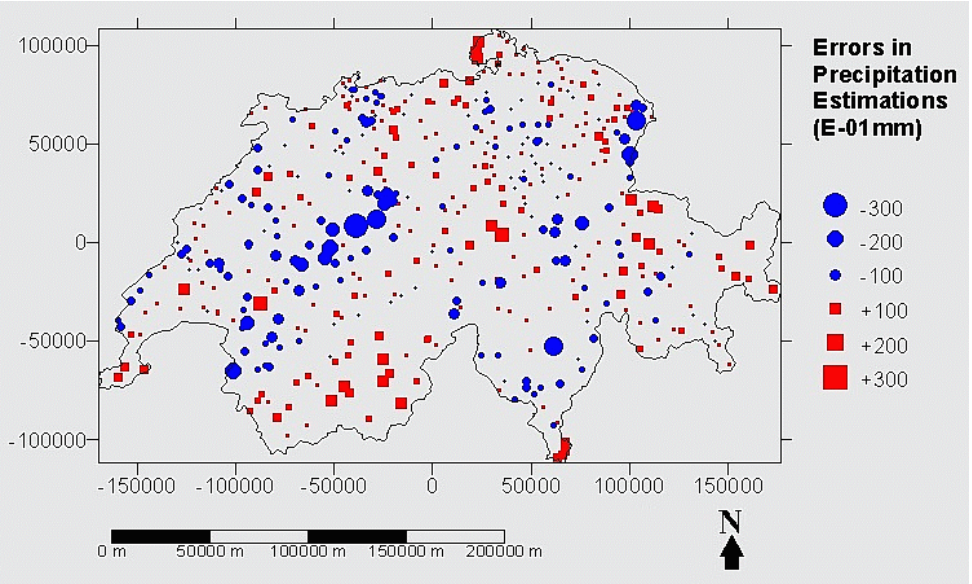

Figure 13 Map of precipitation-estimation errors (estimated value – true value). Symbol radius is proportional to the error.

It should be noticed that, strictly speaking, if we take the trend isolated and modelled in Section 2.3. as our *definition* of the large-scale variation (as Cressie does in his

median-polish kriging (1993)), this trend itself will theoretically never be misleading. The only possible errors rely in the assumptions we make about the residual behaviour.

Figure 14 shows locations of significant errors, with respect to a contour map of the predicted precipitation values. These locations are not random : they are mostly places where the model fit would be expected to be poorer. Such locations are of two types : (1) external areas, where neighbouring data are missing in some directions (definition of large scale variation along at least one of the main directions was not "helped" by the presence values from a contiguous interval, see discussion in Section 2.4.); (2) areas with average precipitation-trend values, but where large-scale variation is fairly quickly changing from high to low, and where slight changes in the parameters defining large-scale variation might lead to significant changes in the local trend value.

There is not much to do about the first of these two location types, except to increase the size of the sampled domain, by adding data from neighbouring countries (or, at least, from stations as close as possible to the country border). Probably another residual variance predictor might be built, taking account of the sparseness of data used to isolate the trend in these areas, but this would have been difficult on the basis of the data set that was used here (too few data points are located in these areas).

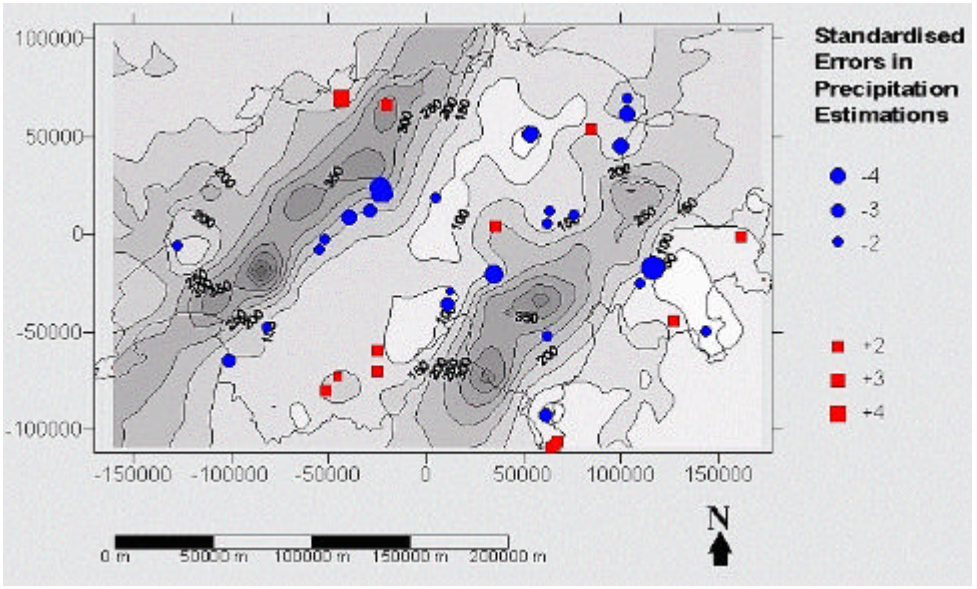

Figure 14 Map showing significant standardised errors ((estimated value – true value) / kriging standard deviation, with absolute values higher than 1.96), with respect to a contour map of estimated precipitation (colour scale, not shown due to lack of space, is as on Figure 6).

The second type of error locations concerns mostly significant underestimation. In fact, we already tried to deal with this problem, by introducing the large-scale-variation gradient into the residual variance predictor (see section 3.1.1), but this correction appears to have been insufficient at several locations. These errors occur quite independently from the actual precipitation value. They are probably mainly due to undersampling (to increase the data-set size with only a few data in these areas would probably lead to a significantly different large-scale variation).

5.3. Prediction of smallest and largest precipitation values

Smallest and largest data are often the most difficult to predict, and this case was no exception. Only one of the ten smallest actual values was amongst the ten smallest predictions of our model. However, we strongly believe that, from the available data set, prediction of at least 8 of these data point is (at best) very hazardous, due to their locations outside the sampled area (4 points in the extreme East, 4 points in the extreme South of the country) : in fact, any better result would probably have to be qualified of luck if not based on meteorological considerations (that were not used here). Nevertheless, our ten lowest predictions fall in the same general areas than the ten smallest actual values.

The method performances were quite better for the ten highest data, of which four are amongst the ten highest predictions. Furthermore, eight of these data fall within the 95%-confidence intervals associated to their prediction under normality assumption. The two main highest-precipitation zones were correctly identified.

6. Applicability of the method and conclusions

The method that was presented in this paper is not highly computing consuming and could, at first glance, easily be automated. Its use in case of an emergency would therefore be quite easy.

Its performances to predict precipitation values within the country (we mean : not to close from the country border and obviously outside the sampled area) appeared quite good. Furthermore, it gave a rapid and quite suitable estimation of the uncertainty associated to the predictions. Ability to predict the lowest values of the data set was fairly small, but this was mainly due to outlying locations of these data. Ability to predict the highest values was quite better. Anyway, the global areas where highest and lowest precipitations occured were correctly identified.

We would like to end with some remarks about the selection of a given method to be used in case of emergency. To be valid on a general basis, choice of a method should be based on comparisons dealing with more than one data set. Precipitation data from different days are likely to behave differently. By the way, a directional two-way mean decomposition might produce less stationary residuals with another data set.

Performances of a method applied to a partial data set are likely to be different from performances of the same method with a larger data set. By taking a sample of the data set, the problem scale is changed, and the boundary between large- and small-scale variation will be modified too. To compare efficiently methods, it would be much more advisable (but without doubt also more difficult) to apply the method with the data density with which it is intended to be routinely used, and to *increase* the data density to test the method performances on the basis of the additional data. In the current case, an optimal sample size could be determined by increasing the sample size up to the moment when adding new data does not modify significantly the large-scale variation structure any more. Again, this should ideally be performed on several data sets from various days, under various meteorological conditions, and results should to be compared.

On a general basis, *interpolation* should never be generalised to *extrapolation* without extreme caution (especially if dealing with any kind of data decomposition). From this point of view, it would be highly preferable, for any country, to obtain data from the closest parts of neighbouring countries, or at least to have as many data as

possible from locations within the country, but next to its borders. In the current example, the extreme East of the country was not represented by *any* data in the interpolated set. On the basis of observed large-scale-variation gradient in several other parts of the couture, the highest precipitation values *might very well* have been there... and would have been absolutely impossible to predict.

Lastly, we would like to notice that extracting the 10 highest predicted values from an interpolation exercise, which usually produces predictions with highly variable associated uncertainty, is not the safest way to decide where to apply radioactivity monitoring in case of nuclear incident... This might be a quite hazardous method for people living in areas where prediction was less extreme with higher uncertainty.

References

Cressie, N. A. C. (1993) *Statistics for Spatial Data*. NewYork: Wiley-Interscience, Wiley Series in Probalility and Statistics.

David, M. (1977) *Geostatistical Ore Reserve Estimation.* Amsterdam: Elsevier.

Deutsch, C. V. and Journel, A. G. (1992) *GSLib, Geostatistical Software Library and User's Guide.* Oxford: Oxford University Press.

Isaaks, E. H. and Srivastava, R. M. (1989) *An Introduction to Applied Geostatistics.* Oxford: Oxford University Press.

Matheron, G. (1962) Traité de Géostatistique appliquée, Tome I. *Mémoires du bureau de Recherches Géologiques et Minières, No.14*. Paris : Editions Technip.

Tukey, **J. W.** (1977) *Exploratory Data Analysis.* Reading, MA : Addison-Wesley.
