

# The Ordering of Spatial Data and the Detection of Multiple Outliers

Andrea CERIOLI and Marco RIANI

In this article we suggest a unified approach to the exploratory analysis of spatial data. Our technique is based on a forward search algorithm that orders the observations from those most in agreement with a specified autocorrelation model to those least in agreement with it. This leads to the identification of spatial outliers—that is, extreme observations with respect to their neighboring values—and of nonstationary pockets. In particular, the focus of our analysis is on spatial prediction models. We show that standard deletion diagnostics for prediction are affected by masking and swamping problems when multiple outliers are present. The effectiveness of the suggested method in detecting masked multiple outliers, and more generally in ordering spatial data, is shown by means of a number of simulated datasets. These examples clearly reveal the power of our method in getting inside the data in a way which is more simple and powerful than it would be using standard diagnostic procedures. Furthermore, the behavior of our algorithm under the null hypothesis of no outliers is investigated through a Monte Carlo experiment. Such simulations are also used to build envelopes for the forward search.

**Key Words:** Forward search; Kriging; Masking; Measurement error; Prediction diagnostic; Simulation envelope.

## 1. INTRODUCTION

Two-dimensional spatial data can be conceived as a realization of the stochastic process

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\}, \quad (1.1)$$

where  $D \subset \mathbb{R}^2$  and  $Z$  has an autocorrelation structure over space (Cressie 1993, p. 8). In this article we consider the case of real-valued spatial data, where a Gaussian assumption for  $Z$  is usually invoked. If  $D$  is a countable subset in  $\mathbb{R}^2$ ,  $Z(\mathbf{s})$  can be modeled either as an autoregressive or a Markov random field. On the contrary, when  $\mathbf{s}$  varies continuously through  $D$ , the kriging model of geostatistics can be adopted. At an exploratory stage of analysis one is usually interested in the detection of nonstationary pockets, influential observations, and (clusters of) outliers. For these purposes simple graphical tools such as

---

Andrea Cerioli is Associate Professor, and Marco Riani is Assistant Professor, Istituto di Statistica, Università di Parma, 43100 Parma, Italy (Email: statec1@ipr.univ.cce.unipr.it).

©1999 American Statistical Association, Institute of Mathematical Statistics,  
and Interface Foundation of North America

*Journal of Computational and Graphical Statistics, Volume 8, Number 2, Pages 239–258*

pocket plots and variogram clouds are now commonly available (see, e.g., Haslett et al. 1991; Panatier 1996). After the model-fitting procedure, formal deletion diagnostics can also be computed (Martin 1992; Christensen, Johnson, and Pearson 1992; Haslett and Hayes 1998).

In a spatial context local anomalies are of paramount importance, and spatial outliers are usually defined as observations that are extreme with respect to their neighboring values. They may be isolated, grouped, or dispersed throughout  $D$ . From both a theoretical and a practical point of view, however, it is difficult to distinguish between a cluster of spatial outliers and a pocket of nonstationary observations. In this article we suggest a unified approach to the exploratory analysis of spatial data. Our technique is based on a forward search algorithm that enables us to detect multiple outliers and/or pockets of nonstationarity without being affected by masking and swamping problems. More generally, the suggested algorithm provides an ordering of the observations according to the proposed model. We can then link the effect of each observation back to features of the original data. The focus of our analysis is on the kriging model, but the proposed method can be conveniently applied to any regression model with spatially autocorrelated errors.

The outline of the article is as follows. In Section 2 we state notation and review some preliminary diagnostics for the kriging model. Furthermore, we show the masking effect exhibited by traditional spatial exploratory techniques when multiple outliers are present in the data. In Section 3 we introduce the forward search algorithm for the ordering of spatial data. We also suggest different approaches to the definition of a suitable “starting point” and alternative methods for progressing in the search. The power of our method in the detection of masked multiple outliers is shown in Section 4 by means of a number of artificial datasets where outliers are known to exist. In addition, the performance of our technique both with well-behaved and contaminated data is assessed through Monte Carlo simulations. In Section 5 we give concluding remarks.

The calculations in this article have been performed using GAUSS (Aptech Systems 1994). The software and the data not reported in the article are available from the authors upon request.

## 2. DIAGNOSTICS FOR THE KRIGING MODEL

Ordinary kriging is a popular method for predicting unobservable quantities from the random field (1.1), when  $D$  is a continuous region in  $\mathbb{R}^2$  and

$$Z(\mathbf{s}) = \mu + \delta(\mathbf{s}) \quad \mathbf{s} \in D. \quad (2.1)$$

In Equation (2.1),  $\mu$  is a fixed but unknown constant and  $\{\delta(\mathbf{s}) : \mathbf{s} \in D\}$  denotes an intrinsically stationary (Gaussian) random field with mean zero. The observed process has variogram  $2\gamma(\mathbf{s} - \mathbf{t}) = \text{var}(Z(\mathbf{s}) - Z(\mathbf{t}))$ ;  $\mathbf{s}, \mathbf{t} \in D$ . Given observations at a network  $S \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  of  $n$  spatial locations, the best linear unbiased predictor of the random variable  $Z(\mathbf{s}_0)$ ,  $\mathbf{s}_0 \in D$ , is  $\hat{Z}(\mathbf{s}_0|S) \equiv \hat{Z}(\mathbf{s}_0|Z(\mathbf{s}_i) : i = 1, \dots, n) = \boldsymbol{\lambda}'\mathbf{Z}$ , where

$$\boldsymbol{\lambda}' = \left( \gamma + \mathbf{1} \frac{\mathbf{1} - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}} \right)' \boldsymbol{\Gamma}^{-1}, \quad (2.2)$$

$\boldsymbol{\gamma}' = [\gamma(\mathbf{s}_0 - \mathbf{s}_1), \dots, \gamma(\mathbf{s}_0 - \mathbf{s}_n)]$ ,  $\mathbf{1}' = [1, \dots, 1]$ ,  $\boldsymbol{\Gamma}$  is the  $n \times n$  matrix whose elements are given by  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ ,  $i, j = 1, \dots, n$ , and  $\mathbf{Z}' = [Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)]$ . If  $\boldsymbol{\Gamma}$  is known, the predictor  $\hat{Z}(\mathbf{s}_0|S)$  is best in the sense that it minimizes the mean-squared prediction error (or kriging variance) among all functions  $\boldsymbol{\beta}'\mathbf{Z}$  such that  $E[\boldsymbol{\beta}'\mathbf{Z}] = E[Z(\mathbf{s}_0)]$ , for  $\boldsymbol{\beta}$  an  $n$ -dimensional vector of constants. Specifically,

$$\sigma^2(\mathbf{s}_0|S) \equiv E[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0|S)]^2 = \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} - \frac{(\mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} - 1)^2}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}}.$$

The kriging predictor has the possibly unappealing feature of providing a perfect fit to the observed data; that is,  $\hat{Z}(\mathbf{s}_0|S) = Z(\mathbf{s}_0)$  when  $\mathbf{s}_0 \in S$  (Cressie 1993, pp. 128–129). However, if measurement error affects subsequent observations at the same location, an extra term is added to the right side of (2.1), so that

$$Z(\mathbf{s}) = \mu + \delta(\mathbf{s}) + \epsilon(\mathbf{s}) \quad \mathbf{s} \in D, \quad (2.3)$$

where  $\epsilon(\mathbf{s})$  denotes a zero-mean, white-noise process, independent of  $\delta(\mathbf{s})$ , whose variance defines the measurement error variability. Interest then lies in predicting the noiseless version  $\delta(\mathbf{s})$  of  $Z(\mathbf{s})$ . In this case, the kriging predictor and its variance are given in Appendix A as a function of  $\text{var}(\epsilon(\mathbf{s}))$ , when  $\delta(\mathbf{s})$  is a second-order stationary process.

Prior to the computation of the optimal weight vector  $\boldsymbol{\lambda}$ , it is important to identify observations that are outlying with respect to the specified null model, either (2.1) or (2.3). As emphasis is mainly on prediction, case-deletion diagnostics for kriging are based on standardized prediction residuals computed by cross-validation. Let  $Z_i \equiv Z(\mathbf{s}_i)$ ,  $S_{-i} \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \dots, \mathbf{s}_n\}$ , and  $\hat{Z}_{i(S_{-i})}$  be the kriging predictor for  $Z_i$  based on observations in  $S_{-i}$ . Furthermore, denote by  $\sigma_{i(S_{-i})}^2$  the corresponding kriging variance. Standardized prediction residuals are then defined as

$$\tilde{\epsilon}_{i(S_{-i})} = \frac{Z_i - \hat{Z}_{i(S_{-i})}}{\sigma_{i(S_{-i})}} \quad i = 1, \dots, n. \quad (2.4)$$

Cressie (1986, 1993) suggested the use of exploratory plots of observed residuals  $\tilde{\epsilon}_{i(S_{-i})}$  for the purpose of detecting spatial outliers. Christensen, Johnson, and Pearson (1992) developed formal deletion diagnostic procedures based on (2.4), both with and without measurement error. They also considered the case of universal kriging, where  $\mu$  is not constant over space. However, it is well known that single-case deletion statistics are prone to masking and swamping when multiple outliers are present in the data (Barnett and Lewis 1994). This is true also for standardized prediction residuals, as appears from the following example.

## 2.1 EXAMPLE 1: MULTIPLE OUTLIERS

In our first example we take a realization from the random field (2.3) at the nodes of a  $9 \times 9$  regular grid ( $n = 81$ ). Sites on the grid are indexed in lexicographical order. Data are simulated assuming a Gaussian distribution for both  $\delta$  and  $\epsilon$ , and a spherical

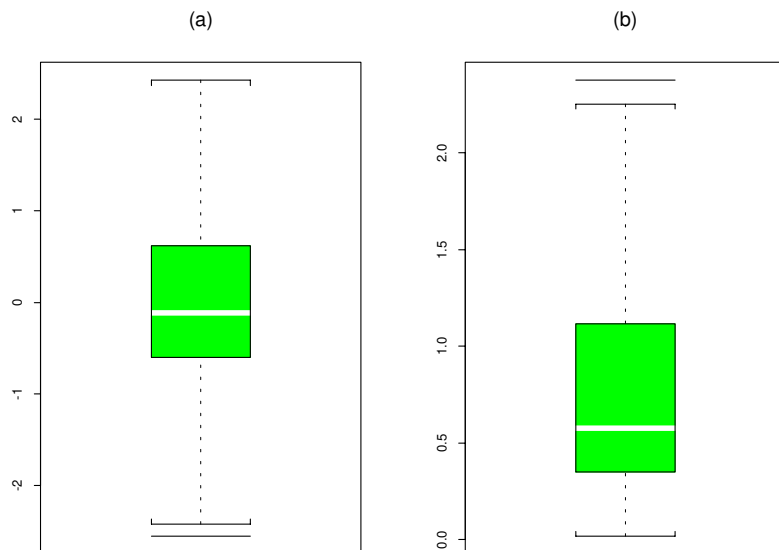


Figure 1. Example 1. Boxplots of standardized prediction residuals (a) and square root of Cook distances (b).

semivariogram

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \|\mathbf{h}\| = 0, \\ \theta_0 + \theta_1[1.5\|\mathbf{h}\|/\theta_2 - .5(\|\mathbf{h}\|/\theta_2)^3] & 0 < \|\mathbf{h}\| \leq \theta_2, \\ \theta_0 + \theta_1 & \|\mathbf{h}\| \geq \theta_2, \end{cases}$$

where  $\mathbf{h} = \mathbf{s} - \mathbf{t}$  and  $\|\cdot\|$  denotes the Euclidean norm. Specifically, we set  $\mu = 10$ ,  $\theta_0 = 2$ ,  $\theta_1 = 4$ , and  $\theta_2 = 8$ ; we also assume that  $\text{var}(\epsilon(s_i)) = .1$  for  $i = 1, \dots, n$ . Simulated data are reported in Appendix B. Then, we modify three values in the northwest corner of the grid, at locations  $s_1$ ,  $s_2$ , and  $s_3$ , respectively, by adding constants 6, 4, and 5 to the original observations. As a result, we obtain contaminated values which are clearly outliers with respect to the bulk of the data.

For such data Figure 1 shows boxplots of both standardized prediction residuals  $\tilde{\epsilon}_{i(S_{-i})}$ , and square root of Cook distances. Under model (2.3), Cook distances for prediction are computed from the estimated semivariogram as

$$C_{i(S_{-i})} = -(\hat{\delta}_{(S_{-i})} - \hat{\delta})' \hat{\Gamma}^{-1} (\hat{\delta}_{(S_{-i})} - \hat{\delta}) \quad i = 1, \dots, n,$$

where  $\hat{\delta}$  denotes the vector of predicted values of  $\delta(s_i)$  based on  $S$ ,  $\hat{\delta}_{(S_{-i})}$  is the corresponding vector of predicted values given  $S_{-i}$ , and computation of  $\hat{\Gamma}$  is discussed in Section 3. None of the contaminated values has prediction diagnostics which can be judged extreme with respect to the specified Gaussian model. Therefore, there is clear evidence of a masking effect due to the presence of multiple outliers. Indeed, the only apparent outlier displayed both in Figure 1(a) and 1(b) is the observation at site  $s_{12}$ , where the relatively low value  $\tilde{\epsilon}_{12(S_{-12})}$  (and hence the high value  $C_{12(S_{-12})}$ ) has been swamped by spatial proximity to the contaminated corner.

As an referee pointed out, such undesirable effects may still be present even if we apply other exploratory techniques specifically devised for locating spatial anomalies.

Table 1. Example 1. Absolute values of the standardized (mean–median) difference  $u$ 

|       | Row number    |     |     |    |    |     |     |     |     |
|-------|---------------|-----|-----|----|----|-----|-----|-----|-----|
|       | 1             | 2   | 3   | 4  | 5  | 6   | 7   | 8   | 9   |
| $ u $ | 1.1           | .6  | 1.0 | .8 | .5 | 1.4 | 4.3 | 3.4 | .5  |
|       | Column number |     |     |    |    |     |     |     |     |
|       | 1             | 2   | 3   | 4  | 5  | 6   | 7   | 8   | 9   |
| $ u $ | 5.1           | 2.3 | .4  | .2 | .4 | 2.0 | 1.0 | 1.8 | 2.1 |

For instance, Table 1 shows for each row and column of the grid the absolute value of the standardized (mean–median) difference (Cressie 1993, p. 38)

$$u \equiv m^{1/2}(\bar{Z} - \tilde{Z})/(.7555\hat{\psi}), \quad (2.5)$$

where  $m$  is the number of spatial sites located on a specific row or column of the grid, and  $\bar{Z}$  and  $\tilde{Z}$  are, respectively, the average and median value computed on that row or column. In addition,  $\hat{\psi}$  is a resistant measure of dispersion computed as

$$\hat{\psi} = (\text{interquartile range})/1.349.$$

Values of  $|u|$  around 3 or greater are usually adopted to highlight atypical rows and columns with gridded data. Thus, in our example, the summary diagnostic picks out rows 7 and 8 and column 1 as having potential outliers. In fact, rows 7 and 8 do not have outliers and only column 1 does. On the contrary, columns 2 and 3 and row 1, which contain outliers, are not chosen by the mean–median summary method.

Of course, some simple diagnostic methods may occasionally provide useful guidance in the identification of multiple outliers. For instance, the use of bivariate plots of  $Z(\mathbf{s}_i)$  and  $Z(\mathbf{s}_i + \mathbf{e})$ , for  $i = 1, \dots, n$  and  $\mathbf{e}$  a unit vector in a cardinal direction (see Cressie 1993, p. 38), can display the three contaminated values in this example for a careful choice of  $\mathbf{e}$ . However, the use of bivariate scatterplots can lead different investigators to different conclusions as to a point outlier status, according to which direction is actually chosen. Thus, we believe that the technique developed in the following sections has the major advantage of making the detection process more formal and, hence, more objective. Furthermore, it provides additional information through the ordering of the observations from those most in agreement with the specified spatial model to those least in agreement with it.

### 3. THE FORWARD SEARCH WITH SPATIALLY AUTOCORRELATED DATA

Our approach to the identification of masked multiple outliers in spatial prediction models is based on a fast forward search through the data. Related algorithms have recently been suggested in regression models with independent errors, multivariate analysis, and data transformations (Hadi 1992; Hadi and Simonoff 1993; Atkinson 1994; Riani and Atkinson 1999). Our proposal extends this field of research to the case of dependent

observations. Furthermore, our method provides an ordering of the data, from closest to the assumed spatial model to furthest from it. This can have beneficial consequences in detecting nonstationary pockets and other potential anomalies at a local scale.

The main steps of our algorithm are as follows.

### 3.1 STEP 0: ESTIMATION OF $\Gamma$

We initially compute a resistant estimate of spatial dependence at lag  $\mathbf{h}$ . In all examples that follow, we adopt the robust variogram estimator of Cressie and Hawkins (1980)

$$2\tilde{\gamma}(\mathbf{h}) = \frac{\left(|N(\mathbf{h})|^{-1} \sum_{N(\mathbf{h})} \sqrt{|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|}\right)^4}{.457 + .494/|N(\mathbf{h})|} \quad \mathbf{h} \in \mathbb{R}^2,$$

where  $|N(\mathbf{h})|$  is the number of pairs of sites at lag  $\mathbf{h}$  and  $\sum_{N(\mathbf{h})}$  denotes summation over such pairs. As an alternative suggestion, Hawkins and Cressie (1984) also considered the use of  $2\tilde{\gamma}^*(\mathbf{h}) = [\tilde{Z}(\mathbf{h})]^4 / .457$ , where  $\tilde{Z}(\mathbf{h})$  denotes the median of square-root differences  $\sqrt{|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|}$  computed over all pairs of sites at lag  $\mathbf{h}$ .

To ensure conditional negative-definiteness of the resulting estimate of  $\Gamma$ , we fit a valid parametric model to sample values  $\tilde{\gamma}(\mathbf{h})$ , leading to model-based estimates  $\hat{\gamma}(\mathbf{h})$  and  $\hat{\Gamma} = [\hat{\gamma}(\mathbf{s}_i - \mathbf{s}_j)]$ . Alternatively, a nonparametric approach to variogram fitting could be taken at this stage (Shapiro and Botha 1991; Barry and ver Hoef 1996). However, as we see in Section 4, results from our search algorithm seem to be largely unaffected by misspecification of the variogram function, when multiple outlier detection and ordering of the data are the focus. For such purposes the computational complexity of fitting general nonparametric families of variogram functions can thus be avoided.

### 3.2 STEP 1: CHOICE OF THE INITIAL SUBSET

The forward search starts with the definition of a subset of  $p$  spatial locations. Observations in this subset are intended to be outlier free. If  $n$  is moderate and  $p \ll n$ , the choice can be performed by exhaustive enumeration of all  $\binom{n}{p}$  distinct  $p$ -tuples  $S_{i_1, \dots, i_p}^{(p)} \equiv \{\mathbf{s}_{i_1}, \dots, \mathbf{s}_{i_p}\}$ , for  $1 \leq i_1, \dots, i_p \leq n$  and  $i_j \neq i_{j'}$ . Specifically, let  $\boldsymbol{\iota}' = [i_1, \dots, i_p]$  and  $\hat{Z}_{i(S_{\boldsymbol{\iota}'})}$  be the kriging predictor at site  $\mathbf{s}_i$  given observations in  $S_{\boldsymbol{\iota}'}$ . With a slight abuse of notation, if  $\mathbf{s}_i \in S_{\boldsymbol{\iota}'}$ , then  $\hat{Z}_{i(S_{\boldsymbol{\iota}'})}$  stands equivalently for the predictor of either the observed process  $Z(\mathbf{s})$  or its noiseless version  $\delta(\mathbf{s})$ , according to the assumed model. The corresponding standardized prediction residual is  $\tilde{e}_{i(S_{\boldsymbol{\iota}'})}$ . We take as our initial subset the  $p$ -tuple  $S_*^{(p)}$  which satisfies

$$\tilde{e}_{[\text{med}](S_*^{(p)})}^2 = \min_{\boldsymbol{\iota}} [\tilde{e}_{[\text{med}](S_{\boldsymbol{\iota}'})}^2], \quad (3.1)$$

where  $\tilde{e}_{[l](S_{\mathbf{l}}^{(p)})}^2$  is the  $l$ th ordered squared residual among  $\tilde{e}_{i(S_{\mathbf{l}}^{(p)})}^2, i = 1, \dots, n$ ,

$$\text{med} = p + \left\lceil \frac{(n-p)}{2} \right\rceil, \tag{3.2}$$

and  $\lceil (n-p)/2 \rceil$  denotes the rounded value of  $(n-p)/2$ . Criterion (3.1) is chosen to provide a robust fit to the data. It is similar to the least median of squares method for regression models with independent errors (Rousseeuw 1984; Atkinson 1986; Hawkins 1993), except that here normalization with respect to  $\sigma_{i(S_{\mathbf{l}}^{(p)})}$  is adopted in the computation of the  $i$ th residual.

In our applications to ordinary kriging, where  $\mu$  is the only large-scale parameter to be estimated from the data, we start from  $p = 2$ , as this is the smallest dimension for which  $\hat{Z}_{i(S_{\mathbf{l}}^{(p)})}$  can be computed. Alternative methods for the definition of an initial subset when  $n$  or  $p$  are large, and exhaustive search of elemental sets becomes infeasible, are discussed in Section 3.5.

### 3.3 STEP 2: PROGRESSING IN THE FORWARD SEARCH

Given a subset of dimension  $m \geq p$ , say  $S_*^{(m)}$ , the forward search moves to dimension  $m + 1$  by selecting the  $m + 1$  spatial locations with the smallest squared standardized prediction residuals, the locations being chosen by ordering all squared residuals  $\tilde{e}_{i(S_*^{(m)})}^2, i = 1, \dots, n$ . It is easy to prove that this criterion extends Hadi and Simonoff (1993, eq. (3)) to the case of spatial prediction models. Specifically, in the context of linear models with independent errors, Hadi and Simonoff based their search on ordering the absolute value of internally studentized residuals for the units belonging to  $S_*^{(m)}$ , and of scaled prediction errors for those not in  $S_*^{(m)}$ .

In most moves from  $m$  to  $m + 1$  just one new site joins the subset. Under model (2.3), it may also happen that two or more locations join  $S_*^{(m)}$  as one or more leave. However, according to our experience, such events are extremely unusual if the algorithm is started from  $S_*^{(p)}$ , and only occur when the search has included one location which belongs to a cluster of spatial outliers. Notice that leaving the subset is not possible under model (2.1), since in that case

$$\tilde{e}_{i(S_*^{(m)})} = 0 \quad \text{if} \quad \mathbf{s}_i \in S_*^{(m)}.$$

A virtually identical search is obtained by ordering the  $n$  squared residuals  $r_{i(S_*^{(m)})}^2$  defined by

$$\begin{aligned} r_{i(S_*^{(m)})}^2 &= e_{i(S_*^{(m)})}^2 & \text{if} & \quad \mathbf{s}_i \in S_*^{(m)}, \\ r_{i(S_*^{(m)})}^2 &= \tilde{e}_{i(S_*^{(m)})}^2 & \text{if} & \quad \mathbf{s}_i \notin S_*^{(m)}, \end{aligned}$$

where  $e_{i(S_*^{(m)})} \equiv Z_i - \hat{Z}_{i(S_*^{(m)})}$ . As in regression models with independent errors (Atkinson and Riani 1997), we have found that using  $r_{i(S_*^{(m)})}^2$  instead of  $\tilde{e}_{i(S_*^{(m)})}^2$  slightly increases the probability that two or more sites join  $S_*^{(m)}$  at the first stages of the forward

search. Therefore, we recommend that this approach be adopted when the initial subset is chosen at random, in order to increase the chance for outliers fortuitously included in the initial subset of being ejected from it (see Section 3.5 and Example 3).

### 3.4 STEP 3: ORDERING SPATIAL DATA

Step 2 of the forward search is repeated until all locations are included in the subset. If just one site enters  $S_*^{(m)}$  at each move, the algorithm provides an ordering of the data according to the specified null model, with observations furthest to it joining the subset at the last stages of the procedure. Spatial outliers and other observations which are potentially anomalous at a local scale can thus be detected by simple graphical displays of a variety of statistics involved in the forward search. In our examples we monitor standardized prediction residuals  $\tilde{\epsilon}_{i(S_*^{(m)})}$ ,  $i = 1, \dots, n$ , for each value of  $m$ , and we draw plots of quantities like

$$\tilde{\epsilon}_m^{(m+1)} = \sqrt{\tilde{\epsilon}_{[m+1](S_*^{(m)})}^2} \quad m = p + 1, \dots, n - 1, \quad (3.3)$$

$$\tilde{\epsilon}_m^{(\max)} = \sqrt{\tilde{\epsilon}_{[n](S_*^{(m)})}^2} \quad m = p + 1, \dots, n - 1, \quad (3.4)$$

and

$$\sigma_m^{(m+1)} = \sigma_{[m+1](S_*^{(m)})}^2 \quad m = p + 1, \dots, n - 1, \quad (3.5)$$

$$\sigma_m^{(\max)} = \sigma_{[n](S_*^{(m)})}^2 \quad m = p + 1, \dots, n - 1. \quad (3.6)$$

If one or more atypical observations are present in the data, the plots of  $\tilde{\epsilon}_m^{(m+1)}$  and  $\sigma_m^{(m+1)}$  must show a peak at the step prior to the inclusion of the first outlier. Correspondingly, with one cluster of outliers the curves of  $\tilde{\epsilon}_m^{(\max)}$  and  $\sigma_m^{(\max)}$  must have a sharp decrease when the first outlier joins  $S_*^{(m)}$ , due to the masking effect. As our examples show, through the joint monitoring of these plots we can then interpret the effect of each observation in terms of features of the original data.

### 3.5 VARIANTS OF THE ALGORITHM

In Step 1 of our algorithm we support the use of exhaustive enumeration of elemental sets to minimize the robust criterion (3.1). However, if  $n$  or  $p$  are large this procedure may require excessive computation. Atkinson (1994) showed how the forward search can be combined with random selection of the initial subset for the purpose of outlier detection. In the case of spatial prediction models, we have tried several criteria for measuring the performance of the  $j$ th search. The best results are obtained through

$$\tilde{p}^{(j)} = \min_m [\tilde{\epsilon}_m^{(\text{med})}], \quad (3.7)$$

where the minimum is taken over all stages of the forward search,

$$\tilde{\epsilon}_m^{(\text{med})} = \sqrt{\tilde{\epsilon}_{[\text{med}](S_*^{(m)})}^2}$$



and med is defined as in (3.2) with  $p$  replaced by  $m$ . Alternative searches from different random starting points are thus compared through quantities  $\tilde{p}^{(j)}$ .

As interest typically lies in the last stages of the forward search, one may wish to initialize the algorithm from  $p^* \gg p$ , especially when  $n$  is very large and running complete searches is computationally costly. In such instances repeated application of the procedure from different random starting points can be adopted, and different choices of  $S_*^{(p^*)}$  compared through values of  $\tilde{p}^{(j)}$ . It is important to remark that, under model (2.3), the assumption that  $S_*^{(p^*)}$  be outlier free is not crucial if  $\text{var}(\epsilon(\cdot))$  is a nonnegligible component of the total variation. In fact, as Example 3 shows, a large amount of measurement error ensures a high degree of interchange among the units belonging to  $S_*^{(p^*)}$  at the first stages of the forward search, and hence makes our technique resistant to the inclusion of some outliers into the initial starting set of locations. For a justification of the importance of measurement error in a proper formulation of kriging see, for example, Cressie (1993, pp. 127–130).

An alternative approximate procedure can be based on the construction of a “pilot fit” of dimension  $p$  and an initial ordering of the data according to such a fit. This is achieved by first selecting  $S_*^{(p)}$  as in Step 1 of the main algorithm. The pilot subset  $S_*^{(p)}$  is then used to find the  $p^*$  spatial locations with the smallest squared standardized prediction residuals

$$\tilde{\epsilon}_{[1](S_*^{(p)})}^2, \quad \dots, \quad \tilde{\epsilon}_{[p^*](S_*^{(p)})}^2. \quad (3.8)$$

Such units form our initial subset of cardinality  $p^*$ . The performance of this approximate search is also assessed in Example 3 in the next section.

## 4. SOME NUMERICAL WORK

In this section we analyze a number of autocorrelated datasets where multiple outliers are known to exist. The power of our approach in detecting masked multiple outliers, and more generally in getting inside the spatial structure of the data, is apparent from our examples irrespective of which variant of the algorithm is actually chosen.

### 4.1 EXAMPLE 1 (CONTINUED): MULTIPLE OUTLIERS

The forward search algorithm based on squared standardized prediction residuals is applied to the simulated data of Example 1, using complete enumeration of all distinct pairs  $\{s_i, s_j\}$  for the selection of  $S_*^{(2)}$ . Figure 2 shows statistics (3.3)–(3.6), restricted for ease of presentation to steps  $m \geq 17$ . Apart from the first stages, where results may be unstable, all such plots lead to the same conclusions and clearly reveal the presence of three outliers. Table 2 reports the units included in the last 10 steps of the forward search for both contaminated and original data. Contaminated locations move into  $S_*^{(m)}$  at the last three steps, with an ordering which reflects their degree of outlyingness.

It should be emphasized that moves at previous steps can also be motivated by inspection of the data given in Appendix B. In fact, they refer to locations whose values are less in agreement with neighboring sites. Such locations are also included at the

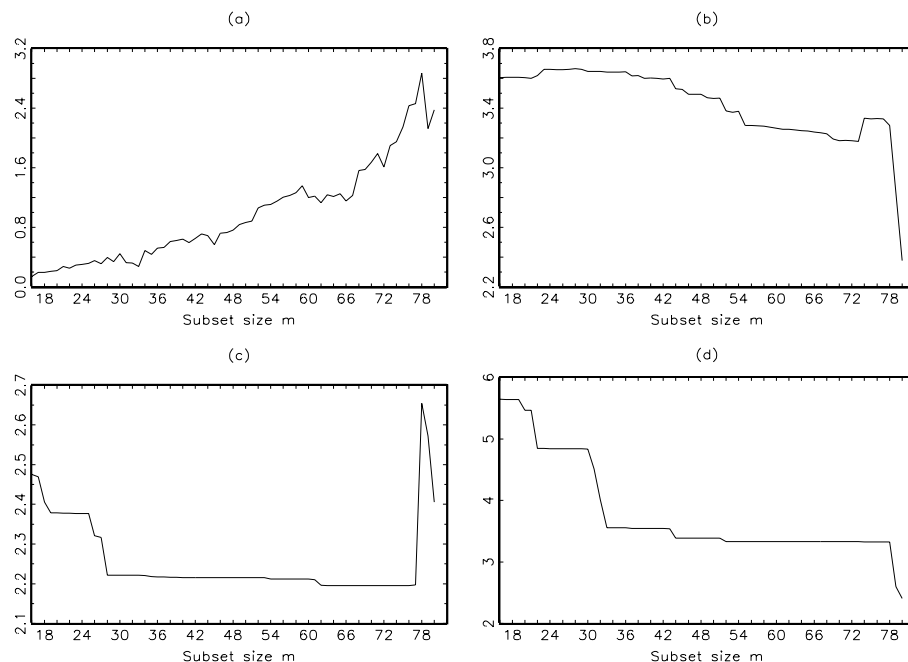


Figure 2. Example 1. Curves of  $\tilde{\epsilon}_m^{(m+1)}$  (a);  $\tilde{\epsilon}_m^{(max)}$  (b);  $\sigma_m^{(m+1)}$  (c); and  $\sigma_m^{(max)}$  (d).

last stages of the forward search run on the original data. Our algorithm thus provides a “natural” ordering of the observations according to the spatial structure implied by model (2.3), even in the case of well-behaved data.

Figure 3 is the (rotated) stalactite plot (Atkinson and Mulira 1993) of standardized prediction residuals  $\tilde{\epsilon}_{i(S_*^{(m)})}$  at successive steps of the forward search. In this plot, \* denotes a value  $|\tilde{\epsilon}_{i(S_*^{(m)})}| > 2.5$ . For ease of presentation we restrict the plot to steps  $m \geq 41$ , and we display only units for which a \* appears in at least one step of the forward search. Furthermore, as  $S_*^{(n)} = S$ , residuals  $\tilde{\epsilon}_{i(S_*^{(n)})}$  must be attributed to measurement error variability and are omitted from the plot. Again, the outlying nature of contaminated observations at locations  $s_1, s_2$ , and  $s_3$  is clear for all subset sizes  $m \leq 78$ , as is the masking effect for  $m > 78$ .

However, the graphical display of the stalactite plot depends on the threshold used. To avoid this choice we suggest to use a plot in which a curve is associated to every

Table 2. Example 1. Units included in the last 10 steps of the forward search

|                   | Steps |       |       |       |       |
|-------------------|-------|-------|-------|-------|-------|
|                   | 71–72 | 72–73 | 73–74 | 74–75 | 75–76 |
| Contaminated data | 18    | 17    | 5     | 34    | 64    |
| Original data     | 12    | 48    | 29    | 17    | 18    |
|                   | 76–77 | 77–78 | 78–79 | 79–80 | 80–81 |
| Contaminated data | 60    | 21    | 1     | 2     | 3     |
| Original data     | 34    | 5     | 64    | 21    | 60    |

|    |   | subset size m |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |
|----|---|---------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|
| s  |   | 41            | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |   |   |   |   |   |   |   |   |   |   |   |   |
| 1  | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * |   |   |   |   |   |   |   |   |
| 2  | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * |   |   |   |   |   |   |
| 3  | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * | * |   |   |   |   |   |
| 17 | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * | * | * |   |   |   |   |
| 21 | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * | * | * | * |   |   |   |
| 34 | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * | * | * | * | * | * |   |
| 60 | * | *             | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | *  | * | * | * | * | * | * | * | * | * | * | * | * |

Figure 3. Example 1. Rotated stalactite plot of standardized prediction residuals. Rows refer to units and columns to steps in the forward search. \* denotes a value  $|\tilde{e}_{i(S_{*}^{(m)})}| > 2.5$ . Units for which  $|\tilde{e}_{i(S_{*}^{(m)})}| \leq 2.5, \forall m \geq 2$ , are omitted from the plot.

unit (Figure 4). Every location is monitored until it joins the subset. In spatial prediction models this plot typically shows a scissors shape, with the sites with small (large) prediction errors joining the subset at the first (last) stages of the forward search. Curves corresponding to atypical observations and to nonstationary pockets stand apart from the others. In addition, relevant features of well-behaved observations can also be grasped from this plot, as appears from the curves associated to locations  $s_{34}$  and  $s_{21}$ , where the smallest and largest uncontaminated values occur.

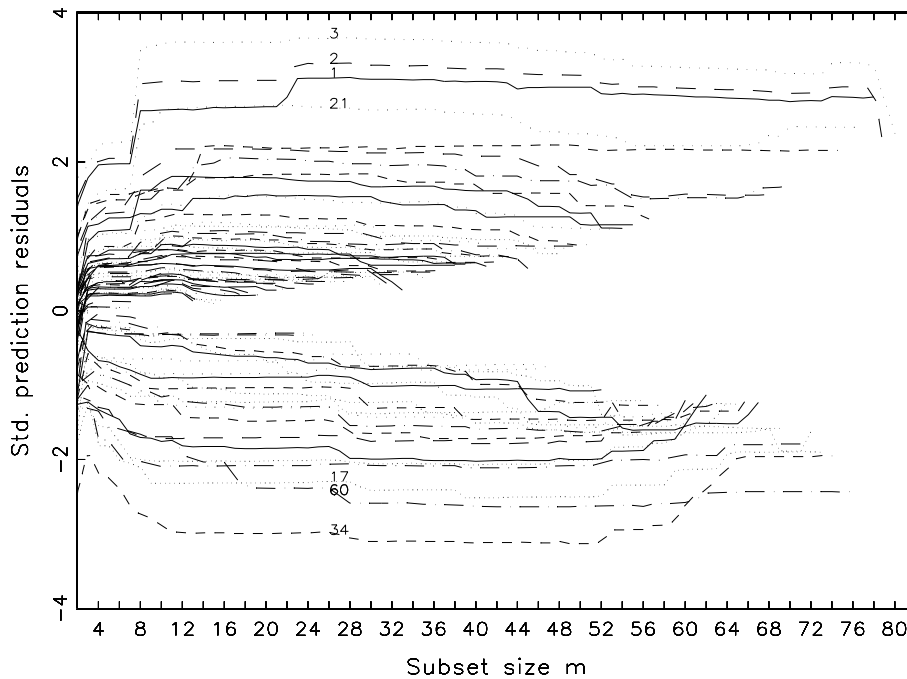


Figure 4. Example 1. Curves of standardized prediction residuals. Each line refers to a unit. Units are monitored up to inclusion in  $S_{*}^{(m)}$ .

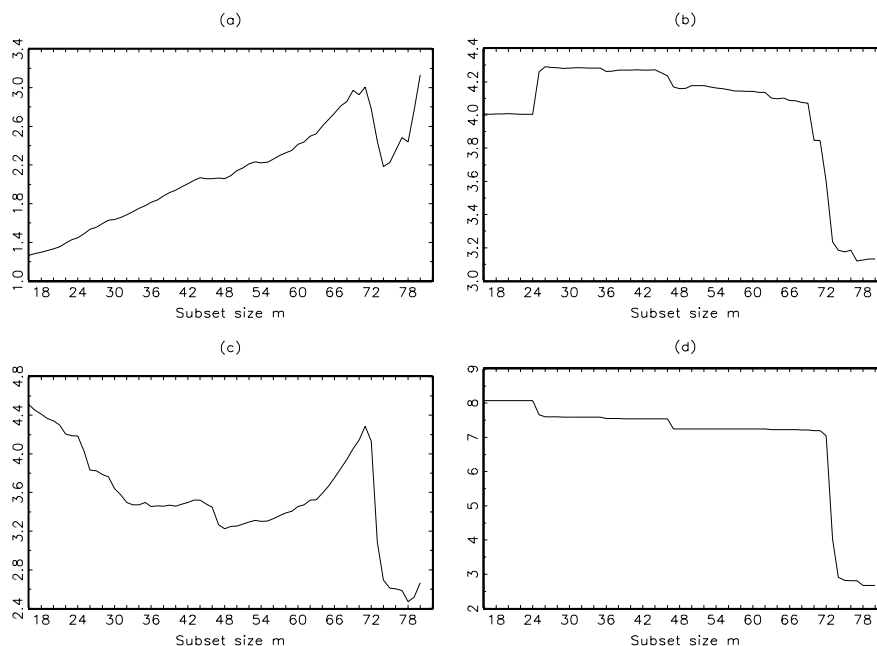


Figure 5. Example 2. Curves of  $\tilde{e}_m^{(ave)}$  (a);  $\tilde{e}_m^{(max)}$  (b);  $\sigma_m^{(ave)}$  (c); and  $\sigma_m^{(max)}$  (d). Estimated spherical semivariogram.

## 4.2 EXAMPLE 2: POCKET OF NONSTATIONARITY

The simulated data in Example 1 are now modified in such a way that  $E[Z(\mathbf{s}_i)] = 18$  for  $i = 1, \dots, 6$ , and  $i = 10, 11, 12$ . Therefore, the problem here is to detect a pocket of nonstationary observations, instead of a cluster of a few spatial outliers. Again, classical spatial exploratory techniques are not able to correctly identify the presence of contaminated values. In fact, in the first two rows of the grid, where the nonstationary area is actually located, the standardized (mean-median) difference (2.5) takes the values  $u = -1.4$  and  $u = -0.2$ , respectively.

Figure 5 displays plots of

$$\tilde{e}_m^{(ave)} = \frac{1}{n-m} \sum_{i=m+1}^n \sqrt{\tilde{e}_{[i](S_*^{(m)})}^2}$$

and

$$\sigma_m^{(ave)} = \frac{1}{n-m} \sum_{i=m+1}^n \sigma_{[i](S_*^{(m)})}^2,$$

as well as  $\tilde{e}_m^{(max)}$  and  $\sigma_m^{(max)}$ , for  $m = 17, \dots, 80$ . There is clear evidence of a different behavior in all curves when passing from  $m = 71$  to  $m = 72$ , and  $\mathbf{s}_5$  enters  $S_*^{(m)}$ . This site corresponds to the smallest contaminated value and is the first location in the

Table 3. Example 2. Units included in the last 12 steps of the forward search

|                              | Steps |       |       |       |       |       |
|------------------------------|-------|-------|-------|-------|-------|-------|
|                              | 69-70 | 70-71 | 71-72 | 72-73 | 73-74 | 74-75 |
| Estimated spherical semivar. | 21    | 64    | 5     | 1     | 3     | 2     |
| Estimated linear semivar.    | 5     | 1     | 3     | 2     | 12    | 21    |
|                              | 75-76 | 76-77 | 77-78 | 78-79 | 79-80 | 80-81 |
| Estimated spherical semivar. | 12    | 4     | 10    | 11    | 60    | 6     |
| Estimated linear semivar.    | 4     | 64    | 11    | 10    | 60    | 6     |

nonstationary pocket that is included by the forward search. Other sites in that area follow at subsequent steps, as is clear from Table 3. Note that the larger contamination in the northwest corner of the grid now worsens the prediction of the low value  $Z(s_{60})$ . Sites  $s_{21}$  and  $s_{64}$ , where the two largest uncontaminated observations occur, move into  $S_*^{(m)}$  at steps  $m = 69$  and  $m = 70$ , respectively, and give rise to the first peak in the plot of  $\tilde{e}_m^{(ave)}$ . Again, the ordering provided by all curves in Figure 5 is consistent with the actual structure of the data, with spatial outliers and observations belonging to the nonstationary pocket included at the last stages of the forward search.

Residuals in Figure 5 are obtained by forcing sample values  $\tilde{\gamma}(\mathbf{h})$  to fit a spherical semivariogram function. However, very similar results are reached even if we dismiss any prior information about the true form of  $\gamma(\mathbf{h})$ , and we fit a simple linear semivariogram in the range  $0 \leq \|\mathbf{h}\| \leq \|s_1 - s_n\|$ . Inspection of the plots based on the linear semivariogram (not reported, but available upon request) shows that the only effect of this misspecification is a slight advance in the abrupt change experienced by all curves, with values  $Z(s_{21})$  and  $Z(s_{64})$  which now would be (wrongly) declared as patent spatial outliers by the forward search. Units included in the last 12 steps of the forward search are also given in Table 3 and are the same as for the true semivariogram function, although in a slightly different order.

### 4.3 EXAMPLE 3: LARGE DATASET

In this example we take a realization from the random field (2.3) at the nodes of a  $20 \times 20$  regular grid. Data are simulated from a Gaussian distribution with  $\mu = 10$  and a spherical semivariogram. Parameter values are set as in Example 1. In addition, we assume here that  $\text{var}(\epsilon(s_i)) = .5$  for  $i = 1, \dots, n$ . Contamination is performed by setting  $E[Z(s_i)] = 22$  at 20 sites on the grid. Spatial coordinates of modified data are displayed in Figure 6, where both clusters of nonstationary observations and isolated spatial outliers appear. A sample size as large as  $n = 400$  might be considered of borderline applicability for the exhaustive enumeration implied by the least median of squares criterion (3.1) and for complete running of the forward search. Therefore, this example is intended to provide a comparison between our standard algorithm and its variants based on selection of an initial subset  $S_*^{(p^*)}$ , for  $p^* \gg p$ , in situations where computing time may become a problem.

As a reference standard, Figures 7(a) and 7(b) respectively show the plots of  $\tilde{e}_m^{(m+1)}$

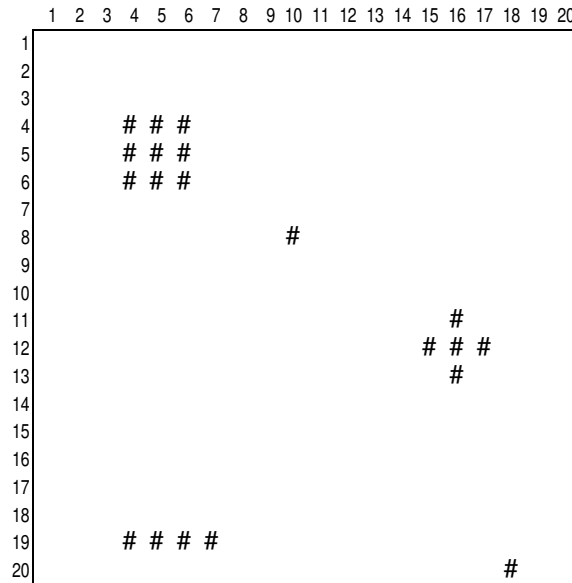


Figure 6. Example 3. Spatial locations (#) of contaminated values on a  $20 \times 20$  regular grid.

and  $\sigma_m^{(\max)}$  for  $m \geq 340$ , when the initial subset  $S_*^{(2)}$  is chosen by exhaustive enumeration of elemental pairs and the search is based on  $\tilde{e}_{i(S_*^{(m)})}^2$ . Instead, Figure 7(c) displays the plot of  $\sigma_m^{(\max)}$  obtained from a pilot fit of dimension  $p = 2$ , as described in Equation (3.8), with  $p^* = 340$ . Figure 7(d) reports the same diagnostic curve for the best forward search out of 25, according to criterion (3.7), when the dimension of the initial subset is  $p^* = 340$ ,  $S_*^{(340)}$  is randomly selected and the search is run using residuals  $r_{i(S_*^{(m)})}^2$  instead of  $\tilde{e}_{i(S_*^{(m)})}^2$ . There is striking agreement between results from alternative searches, with an abrupt change in all plots when passing from  $m = 380$  to  $m = 381$ . At that step, the first contaminated location—belonging to the nonstationary pocket in the northwest corner of the grid—is included in the subset. Inclusions of contaminated sites from different pockets give rise to the subsequent peaks in the curve of  $\tilde{e}_m^{(m+1)}$ , and to the corresponding elbows in those of  $\sigma_m^{(\max)}$ .

Being initialized at  $p = 2$ , our main algorithm provides more stable results than its approximate variants at the first displayed steps of the forward search. Just a few steps after  $p^*$ , however, the curves plotted in panels (c) and (d) of Figure 7 are practically indistinguishable from that of panel (b). The only perceivable difference between panels (c) and (d) also appears in the first stages, where random selection produces a less smooth behavior of diagnostic summaries. This is a consequence of the large number of badly predicted spatial locations leaving the subset at such stages. In fact, random selection does not ensure that  $S_*^{(p^*)}$  be outlier free. Nevertheless, contaminated values rapidly pull out from the subset if measurement error variability is sufficiently high, as is in the present example.

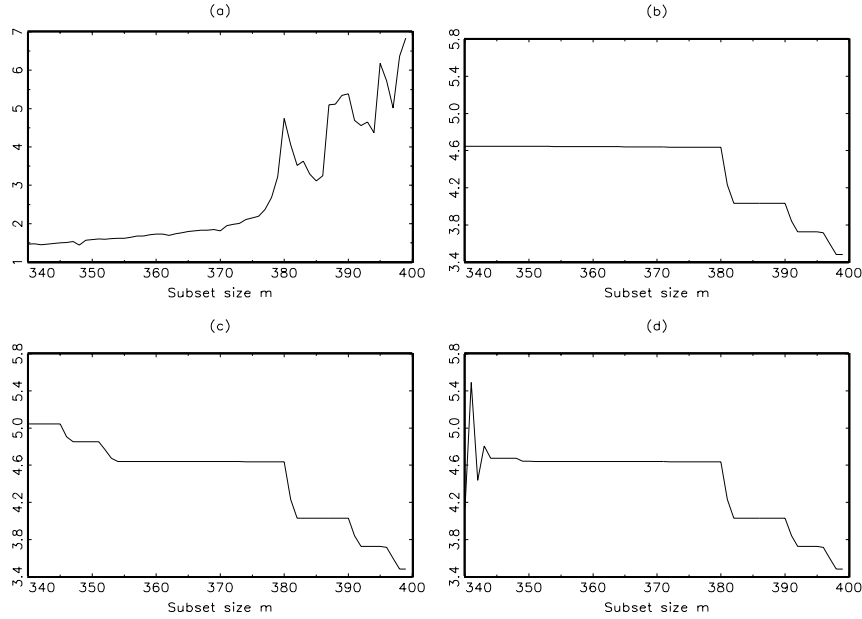


Figure 7. Example 3. Curves of  $\hat{\epsilon}_m^{(m+1)}$  (a) and  $\sigma_m^{(max)}$  (b) for the forward search based on exhaustive enumeration of elemental pairs. Curve of  $\sigma_m^{(max)}$  for the forward search based on a pilot fit of dimension  $p = 2$ , with  $p^* = 340$  (c). Curve of  $\sigma_m^{(max)}$  for the best solution out of 25 randomly started forward searches, with  $p^* = 340$  (d).

Standardized prediction residuals of units not belonging to  $S_*^{(m)}$ ,  $m = p^*, \dots, n - 1$ , are monitored in Figure 8 for both modified versions of our algorithm. With random selection, the ejection of outliers from the initial subset is clearly apparent. After a few steps, all curves corresponding to contaminated sites are practically indistinguishable from those obtained with purposive selection, thus showing that our randomly-started procedure is resistant to the inclusion of some outliers into  $S_*^{(p^*)}$ . From both panels in Figure 8 one can also see that locations belonging to a common nonstationary area (which are represented by lines of the same type) have residual curves of very similar shape.

#### 4.4 MONTE CARLO SIMULATIONS

A small Monte Carlo experiment was carried out in order to investigate the behavior of our algorithm both under the null and the alternative hypothesis. In the first instance, we performed repeated simulation of outlier-free datasets, according to model (2.3) and various assumptions on measurement error variability. As in Example 1, data from a Gaussian distribution with  $\mu = 10$  were generated at the nodes of a  $9 \times 9$  regular grid, following a spherical semivariogram model with  $\theta_0 = 2$ ,  $\theta_1 = 4$ , and  $\theta_2 = 8$ . The forward search algorithm was then run on each simulated data set starting from  $p = 2$ . Figure 9 displays Monte Carlo estimates of the number of potential outliers at successive steps of the forward procedure, when  $\text{var}(\epsilon(s_i)) = .1$  and  $\text{var}(\epsilon(s_i)) = .5$ . Specifically,  $Z(s_i)$

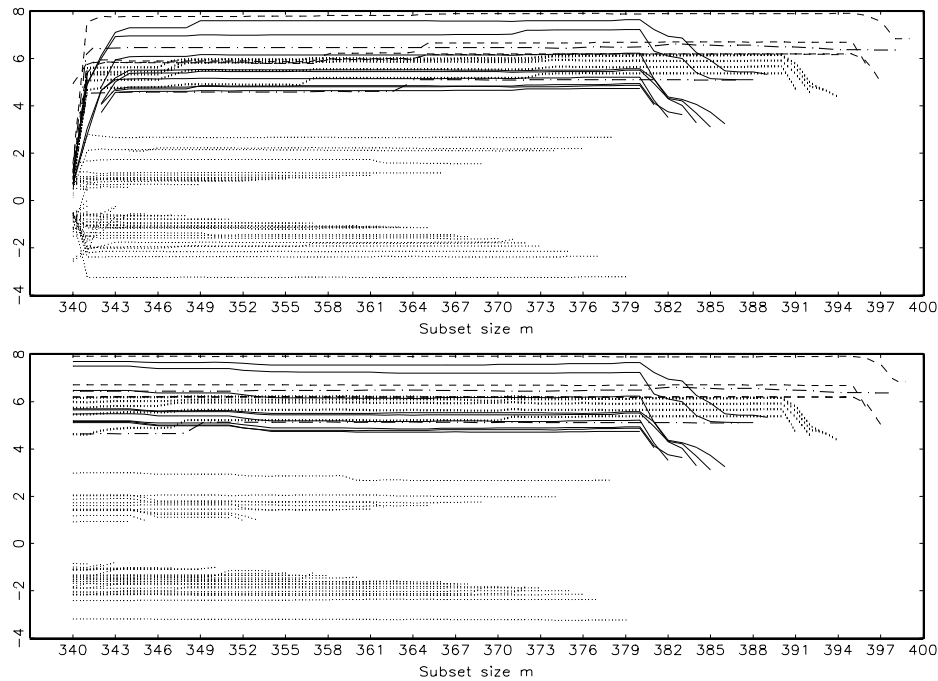


Figure 8. Example 3. Curves of standardized prediction residuals for units not belonging to  $S_*^{(m)}$ . Solid and dashed lines: contaminated sites. Dotted lines: uncontaminated sites. The upper panel refers to the best solution out of 25 randomly started forward searches, with  $p^* = 340$ . The lower panel refers to the search based on a pilot fit of dimension  $p = 2$ , with  $p^* = 340$ . Units are monitored up to their final inclusion in  $S_*^{(m)}$ .

is identified as a potential outlier at subset size  $m$  if  $|\tilde{e}_{i(S_*^{(m)})}| > \alpha$ , for  $\alpha = 2.5$  and  $\alpha = 3$ . Figure 9 reports results based on 500 independent simulations. Such curves do not reveal the method producing spuriously large numbers of outliers, especially at the last steps of the forward search, although there is some evidence that higher cutoffs should be considered when  $\text{var}(\epsilon(s_i))$  is large. Indeed, it must be remarked that an increase in measurement error variability causes the predictor to be less influenced by the correlation structure in the data, as appears from the results in Appendix A. Misspecification of the true variogram had only a minor impact on the good null properties of our algorithm, as very similar results were also obtained by forcing the simulated data to fit a simple linear variogram function.

Monte Carlo simulations of well-behaved data are also used to build confidence bands for statistics obtained from the forward search. Here, 500 independent realizations of model (2.3) are generated setting parameter values equal to their (resistant) estimates of Example 1. Figure 10 shows two-sided 90% simulation envelopes, together with curves of  $\tilde{e}_m^{(m+1)}$  and  $\sigma_m^{(m+1)}$  from Figure 2. Peaks already detected lie above the envelope, thus providing a confirmation of the graphical analyses of Section 4.1.

The performance of our standard algorithm was also investigated under the alternative hypothesis of multiple outliers. For this purpose, 500 independent datasets were again generated as in Example 1. However, contamination is now performed by replac-



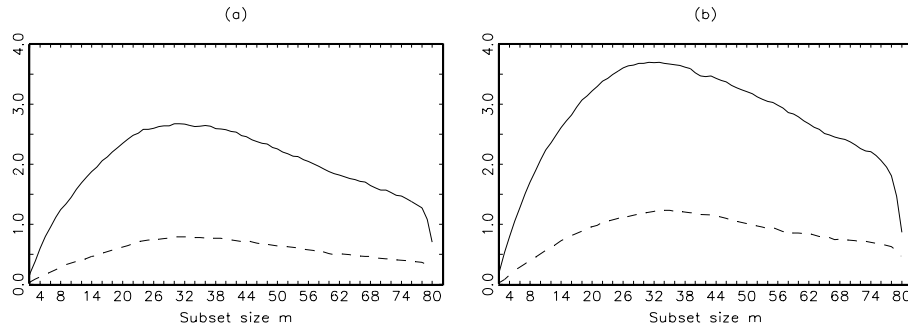


Figure 9. Monte Carlo estimates of the number of sites such that  $|\bar{\epsilon}_{i(S_*^{(m)})}| > \alpha$ , for  $\text{var}(\epsilon(\mathbf{s}_i)) = .1$  (a); and  $\text{var}(\epsilon(\mathbf{s}_i)) = .5$  (b). Solid line:  $\alpha = 2.5$ . Dashed line:  $\alpha = 3.0$ .

ing the smallest  $Z(\mathbf{s}_i)$  and two of its adjacent values with the largest  $Z(\mathbf{s}_i)$  and the corresponding adjacent values. This creates three spatial outliers which are not, in the absence of neighborhood information, disproportionately large values. Furthermore, this variable-site contamination scheme does not rely on the specific choice of a contaminated area in the grid, and hence is less sensitive to where spatial outliers are actually located. Figure 11(a) shows, for each site, Monte Carlo averages of standardized prediction residuals (2.4) at successive steps of the forward search, separately for simulations where the corresponding site has been contaminated (dashed lines) and for simulations where it has not (solid lines). Figure 11(b) displays Monte Carlo estimates (over all simulations) of average prediction residuals at successive steps of the search, both for contaminated values (dashed line) and for uncontaminated observations (solid line). From both panels the effect of contamination is clear, as is the masking effect for  $m > 78$ , even in this situation where spatial outliers are randomly located and their values do not disagree with the bulk of the data. On the contrary, standardized prediction residuals of well-behaved spatial locations are always centered around zero, with a range which is essentially a function of the magnitude of  $\text{var}(\epsilon(\mathbf{s}_i))$ .

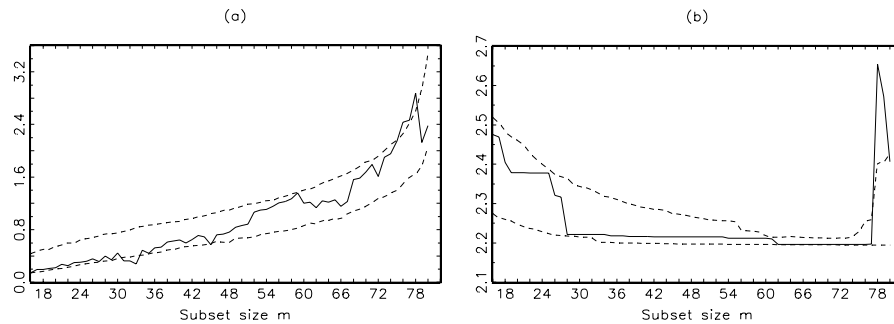


Figure 10. Example 1. 90% envelopes (dashed lines) for  $\epsilon_m^{(m+1)}$  (a), and  $\sigma_m^{(m+1)}$  (b).

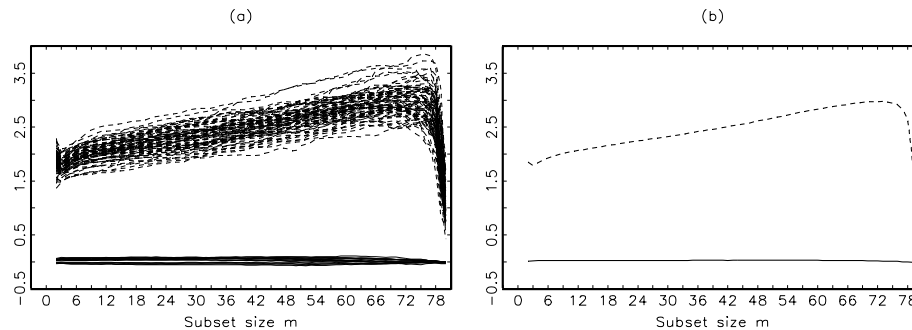


Figure 11. (a) Monte Carlo average of standardized prediction residuals for each site. Dashed lines: simulations where the site has been contaminated. Solid lines: simulations where the site has not been contaminated. (b) Monte Carlo estimate of the average of standardized prediction residuals. Dashed line: contaminated sites. Solid line: uncontaminated sites.

## 5. DISCUSSION

In this article we propose a unified approach to the exploratory analysis of spatial data, which rests upon running a fast forward search algorithm through the data. Our method proves useful in detecting masked multiple outliers, nonstationary pockets and other anomalies at a local scale. Furthermore, it carries the important bonus of ordering the observations with respect to the specified spatial model, with values furthest from it being included at the last steps of the forward search. Therefore, we argue that our approach can convey effective information even in the plain situation where no outlier is present in the data. Results are displayed through simple plots of a variety of statistics monitored along the forward algorithm. These graphical displays are both powerful in revealing the structure of the data and easy to interpret. Our experience shows that deeper understanding of the data can be gathered by joint presentation of multiple plots. Confirmatory analysis is also available via Monte Carlo simulation from the null model, although the emphasis here is not on formal tests of outlyingness.

Our procedure has been successfully applied to the kriging model of geostatistics, where standard diagnostics can suffer from the problems of masking and swamping. The same limitations also affect the robust technique of Hawkins and Cressie (1984), which is only proof against isolated outliers. A resistant estimate of spatial dependence is computed at a preliminary step, using all the data. Nevertheless, it must be remarked that results from our algorithm seem to be largely unaffected by misspecification of the autocorrelation structure in the data.

High-breakdown methods which are based on very robust parameter estimation have been developed for the purpose of identifying masked multiple outliers in regression models with independent errors (see, e.g., Atkinson 1986; Fung 1993; Hawkins and McLachlan 1997). Similar techniques are not available in the case of spatially autocorrelated observations, however. In addition, the interdependencies among spatially autocorrelated data make the use of traditional deletion techniques questionable. Therefore, we consider the forward search algorithm proposed in the present article as a powerful and practical alternative to the development of high-breakdown methods for spatial statistical models.

As a final remark, we note that even if the focus of our analysis is on kriging,

the proposed method can be conveniently applied to any (Gaussian) regression model with spatially autocorrelated errors. In spatial statistics an important goal is to study the effect of removing major assumptions, such as Gaussianity (see, e.g., Diggle, Tawn, and Moyeed 1998), stationarity and fixed spatial locations. How the forward search can cope with these problems will be the subject of further research.

### APPENDIX A: KRIGING PREDICTION WITH MEASUREMENT ERROR

If the process is second-order stationary, it is simpler to write equation (2.2) as a function of  $\mathbf{c}$  and  $\Sigma$ , where  $\mathbf{c}' = [c(s_0 - s_1), \dots, c(s_0 - s_n)]$ ,  $\Sigma$  is the  $n \times n$  covariance matrix whose elements are given by  $c(s_i - s_j)$ ,  $i, j = 1, \dots, n$ , and  $c(\mathbf{s} - \mathbf{t}) = \text{cov}(Z(\mathbf{s}), Z(\mathbf{t}))$ ;  $\mathbf{s}, \mathbf{t} \in D$ . Let equation (2.3) hold with  $\text{var}(\epsilon(\mathbf{s})) = k > 0$ . Then, it can be shown that

$$\hat{\delta}(s_0|S) = k^{-1}\mathbf{c}'\mathbf{Z} - k^{-2}\mathbf{c}'\mathbf{A}\mathbf{Z} + \frac{\mathbf{1}'\mathbf{Z} - k^{-1}\mathbf{1}'(\mathbf{c}\mathbf{1}' + \mathbf{A})\mathbf{Z} + k^{-2}\mathbf{1}'(\mathbf{A}\mathbf{c}\mathbf{1}' + \mathbf{c}\mathbf{1}'\mathbf{A})\mathbf{Z} - k^{-3}\mathbf{1}'\mathbf{A}\mathbf{c}\mathbf{1}'\mathbf{A}\mathbf{Z}}{n - k^{-1}\mathbf{1}'\mathbf{A}\mathbf{1}},$$

where  $\mathbf{1}' = [1, \dots, 1]$ ,  $\mathbf{Z}' = [Z_1, \dots, Z_n]$ ,  $\mathbf{A} = (\Sigma_\delta^{-1} + k^{-1}\mathbf{I}_n)^{-1}$ ,  $\Sigma_\delta = [\text{cov}(\delta(\mathbf{s}_i), \delta(\mathbf{s}_j))]$ ,  $i, j = 1, \dots, n$ , and  $\mathbf{I}_n$  is the  $n \times n$  identity matrix. The corresponding mean-squared prediction error is

$$\sigma_\delta^2(s_0|S) \equiv E[\delta(s_0) - \hat{\delta}(s_0|S)]^2 = k + \sigma_\delta^2 - k^{-1}\mathbf{c}'(\mathbf{I}_n - k^{-1}\mathbf{A})\mathbf{c} + \frac{k - 2\mathbf{1}'\mathbf{c} + 2k^{-1}\mathbf{1}'\mathbf{A}\mathbf{c} + k^{-1}(\mathbf{1}'\mathbf{c} - k^{-1}\mathbf{1}'\mathbf{A}\mathbf{c})^2}{n - k^{-1}\mathbf{1}'\mathbf{A}\mathbf{1}},$$

where  $\sigma_\delta^2 = \text{var}(\delta(\mathbf{s}))$ .

### APPENDIX B: SIMULATED VALUES IN EXAMPLE 1

Spatial locations are at the nodes of a  $9 \times 9$  regular grid.

|   | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    |
|---|------|------|------|------|------|------|------|------|------|
| 1 | 11.4 | 13.4 | 12.8 | 13.2 | 9.85 | 11.4 | 12.0 | 10.8 | 10.5 |
| 2 | 11.8 | 12.7 | 10.8 | 13.8 | 14.4 | 11.3 | 8.73 | 6.79 | 7.21 |
| 3 | 11.7 | 12.3 | 15.6 | 12.3 | 12.3 | 10.2 | 8.95 | 7.68 | 10.6 |
| 4 | 12.8 | 9.95 | 12.8 | 9.70 | 10.4 | 8.36 | 5.46 | 7.19 | 10.0 |
| 5 | 11.6 | 11.9 | 13.8 | 9.98 | 8.89 | 8.46 | 7.53 | 10.6 | 9.11 |
| 6 | 11.5 | 11.4 | 14.5 | 12.6 | 11.6 | 11.6 | 8.35 | 8.69 | 10.4 |
| 7 | 11.4 | 13.1 | 11.3 | 12.4 | 11.3 | 7.43 | 11.0 | 11.4 | 9.49 |
| 8 | 15.6 | 12.3 | 12.2 | 14.0 | 11.6 | 11.4 | 10.8 | 11.5 | 10.2 |
| 9 | 12.7 | 12.6 | 13.8 | 15.0 | 12.8 | 11.6 | 12.2 | 10.9 | 10.8 |

### ACKNOWLEDGMENTS

The authors are grateful to the associate editor, two referees, and Anthony C. Atkinson for valuable comments and suggestions, and to Sergio Zani for his continuing support. This research was supported by the Italian Ministry for Universities and Research.

[Received January 1998. Revised September 1998.]

## REFERENCES

- Aptech Systems, Inc. (1994), *GAUSS*, Maple Valley, WA.
- Atkinson, A. C. (1986), "Masking Unmasked," *Biometrika*, 73, 533–541.
- (1994), "Fast Very Robust Methods for the Detection of Multiple Outliers," *Journal of the American Statistical Association*, 89, 1329–1339.
- Atkinson, A. C., and Mulira, H.-M. (1993), "The Stalactite Plot for the Detection of Multivariate Outliers," *Statistics and Computing*, 3, 27–35.
- Atkinson, A. C., and Riani, M. (1997), "Bivariate Boxplots, Multiple Outliers, Multivariate Transformations and Discriminant Analysis: the 1997 Hunter Lecture," *Environmetrics*, 8, 583–602.
- Barnett, V., and Lewis, T. (1994), *Outliers in Statistical Data* (3rd ed.), New York: Wiley.
- Barry, R. P., and ver Hoef, J. M. (1996), "Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.
- Christensen, R., Johnson, W., and Pearson, L. M. (1992), "Prediction Diagnostics for Spatial Linear Models," *Biometrika*, 79, 583–591.
- Cressie, N. (1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association*, 81, 625–634.
- (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley.
- Cressie, N., and Hawkins, D. M. (1980), "Robust Estimation of the Variogram: I," *Mathematical Geology*, 12, 115–125.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350.
- Fung, W.-K. (1993), "Unmasking Outliers and Leverage Points: A Confirmation," *Journal of the American Statistical Association*, 88, 515–519.
- Hadi, A. S. (1992), "Identifying Multiple Outliers in Multivariate Data," *Journal of the Royal Statistical Society, Ser. B*, 54, 761–771.
- Hadi, A. S., and Simonoff, J. S. (1993), "Procedures for the Identification of Multiple Outliers in Linear Models," *Journal of the American Statistical Association*, 88, 1264–1272.
- Haslett, J., Bradley, R., Craig, P., Unwin, A., and Wills, G. (1991), "Dynamic Graphics for Exploring Spatial Data With Application to Locating Global and Local Anomalies," *The American Statistician*, 45, 234–242.
- Haslett, J., and Hayes, K. (1998), "Residuals for the Linear Model with General Covariance Structure," *Journal of the Royal Statistical Society, Ser. B*, 60, 201–215.
- Hawkins, D. M. (1993), "The Accuracy of Elemental Sets Approximations for Regression," *Journal of the American Statistical Association*, 88, 580–589.
- Hawkins, D. M., and Cressie, N. (1984), "Robust Kriging—A Proposal," *Mathematical Geology*, 16, 3–18.
- Hawkins, D. M., and McLachlan, G. J. (1997), "High-Breakdown Linear Discriminant Analysis," *Journal of the American Statistical Association*, 92, 136–143.
- Martin, R. J. (1992), "Leverage, Influence and Residuals in Regression Models when Observations are Correlated," *Communications in Statistics, Part A—Theory and Methods*, 21, 1183–1212.
- Panatiar, Y. (1996), *Variowin. Software for Spatial Data Analysis in 2D*, New York: Springer-Verlag.
- Riani, M., and Atkinson, A. C. (1999), "A Unified Approach to Multivariate Transformations and Multiple Outliers," submitted for publication.
- Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871–880.
- Shapiro, A., and Botha, J. D. (1991), "Variogram Fitting With a General Class of Conditionally Nonnegative Definite Functions," *Computational Statistics and Data Analysis*, 11, 87–96.