

## **Robust methods for the analysis of spatially autocorrelated data**

**Andrea Cerioli, Marco Riani**

Dipartimento di Economia, Sezione di Statistica e Informatica, Università di Parma, Via Kennedy 6, 43100 Parma, Italy (e-mail: andrea.cerioli@unipr.it, mriani@unipr.it)

**Abstract.** In this paper we propose a new robust technique for the analysis of spatial data through simultaneous autoregressive (SAR) models, which extends the Forward Search approach of Cerioli and Riani (1999) and Atkinson and Riani (2000). Our algorithm starts from a subset of outlier-free observations and then selects additional observations according to their degree of agreement with the postulated model. A number of useful diagnostics which are monitored along the search help to identify masked spatial outliers and high leverage sites. In contrast to other robust techniques, our method is particularly suited for the analysis of complex multidimensional systems since each step is performed through statistically and computationally efficient procedures, such as maximum likelihood. The main contribution of this paper is the development of joint robust estimation of both trend and autocorrelation parameters in spatial linear models. For this purpose we suggest a novel definition of the elemental sets of the Forward Search, which relies on blocks of contiguous spatial locations.

**Key words:** Block forward search, Masking, SAR model, Spatial outliers

### **1 Introduction**

The Forward Search (FS, for short) is a newly developed tool for robust regression analysis and robust multivariate estimation of location and shape. Its origin stems from the work of Hadi (1992), Hadi and Simonoff (1993) and Atkinson (1994) although it is only recently that the great potential of monitoring diagnostics along the search has been clearly recognized and exploited (Cerioli and Riani, 1999; Atkinson and Riani, 2000). The basic ingredients of the FS are a robust start from an outlier-free subset of observations, a criterion for progressing in the search, which allows the subset to increase by one or more observations at each step, and a set of diagnostic tools that are monitored along the search. The robustness of

the FS stems from the very definition of its algorithm, starting from “good” data points and including outliers only at the end of the procedure. Computation of high-breakdown estimators is not required, except possibly at the starting stage. Indeed, the application of efficient likelihood or moment-based methods at subsequent steps of the FS provides the analyst with more powerful tools than those obtained via traditional high-breakdown estimation (e.g. Rousseeuw, 1984; Rousseeuw and van Zomeren, 1990). In addition, it has been shown in many examples that the robustness properties of the FS can be preserved even if the algorithm is initialized from a contaminated subset, thanks to its ability to identify outliers at the outset and replace them with “good” observations (Atkinson et al., 2003).

The flexibility of the FS makes this procedure suitable for extensions to research areas other than multiple regression and multivariate estimation. This is especially true in the study of complex multidimensional systems, where high-breakdown estimators are often difficult to find and unacceptably inefficient. Spatial data analysis is one such field, due to the difficulties raised by having a vector-valued index attached to each observation, by typical clustering of high and low values within the study region, and by the need of jointly estimating variable relationships and spatial autocorrelation.

In a recent article (Cerioli and Riani, 1999) addressed the problem of outlier detection through the FS in the kriging model of geostatistics (Cressie, 1993), where observations are conceived as a partial realization of a continuous-index spatial stochastic process. In kriging it is also possible to compute a robust measure of the spatial dependence structure (i.e. the variogram) based on all the data, which is an advantage for the FS. The FS for kriging of Cerioli and Riani (1999) was run conditionally on this robust variogram estimate.

In the present paper we focus on a class of popular spatial autoregressive models, namely the class of Simultaneous Autoregressive models (SAR, for short), which has been widely adopted for the analysis of lattice data in the last 25 years. Fitting a SAR model sets new and intriguing problems which can not be answered by the FS approach of Cerioli and Riani (1999). The crucial distinction between application of the FS to kriging and to spatial autoregression is that, under the SAR model, a robust estimator of spatial autoregression parameters based on all the data does not exist. Hence, in the FS that we consider in this paper, spatial autoregression parameters must be estimated consistently and efficiently at each step of the algorithm, together with trend parameters. This makes the FS algorithm for SAR models considerably more complex than the one suggested for kriging.

To define an appropriate FS approach for the SAR model, we argue that the basic algorithm for regression models with independent errors (Atkinson and Riani, 2000) must be suitably modified and run over blocks of contiguous spatial locations. We call the resulting algorithm the *Block Forward Search* (BFS, for short). The proposal of a BFS algorithm for SAR models is the main contribution of this paper.

Blocks in the BFS algorithm are defined so as to retain the same dependence structure as the original data set. Hence an analogy might be drawn between the BFS approach of this paper and subsampling techniques for spatially dependent observations (Sherman, 1996; Politis et al., 1999; Heagerty and Lumley, 2000), which also use blocks of contiguous locations to obtain replicates of the statistics

of interest. However, the final decision about the actual block size and the degree of overlapping between different blocks might vary according to the specific goal of the analysis. In the BFS algorithm, where robust estimation is the main purpose, a larger number of relatively small (nonoverlapping) blocks is likely to be preferred, as this choice enhances the finite sample effectiveness of the method.

The potential of our robust technique for the analysis of spatial data is shown through application to both simulated and real data sets, reproducing a number of situations of practical interest. In particular, we address two important issues, the detection of multiple spatial outliers and the detection of high leverage locations. In both cases we show that our approach clearly outperforms standard statistical diagnostic techniques, which are affected by masking and swamping problems. We also analyse the wheat-yield data set of Mercer and Hall (see Cressie, 1993, p. 455), which has been studied widely in the spatial statistical literature. In this real data example we are thus able to compare our findings with results obtained through different methods, and see that spurious information is not introduced in the absence of multiple spatial outliers. In all cases we show that the BFS algorithm provides plenty of information, being supplemented by many clear and effective graphical displays.

The outline of the paper is as follows. In Sect. 2 we briefly introduce the SAR model and we describe some simple diagnostics that are commonly computed to identify spatial outliers and influential observations. We also display the effect of masking on these diagnostics through some simulated examples. The BFS algorithm for spatial autoregression is proposed in Sect. 3. Section 4 shows the practical advantages of the application of the BFS algorithm to the simulated examples and to the Mercer and Hall wheat-yield data set. Concluding remarks and directions for future research are provided in Sect. 5.

All the data sets analysed in this paper are available at the web page <http://www.riani.it/jiss02>.

## 2 Spatial autoregression

### 2.1 The SAR model

Let  $S \equiv \{s_1, \dots, s_n\}$  be a collection of  $n$  spatial sites and  $Y_i \equiv Y(s_i)$  be a random variable observed at site  $s_i$ ,  $i = 1, \dots, n$ . In practice,  $S$  may denote either a set of administrative units (i.e. counties, municipalities, etc.) or the tiles of a regular grid. Spatial relationships between pairs of locations are represented through the simple weighting scheme:

$$\begin{aligned} w_{ij} &= 1 && \text{if sites } s_i \text{ and } s_j \text{ are neighbours,} \\ w_{ij} &= 0 && \text{otherwise,} \end{aligned}$$

and  $w_{ii} = 0$ . For a regular grid the most common definition of a neighbourhood structure is that for which  $w_{ij} = 1$  if  $s_j$  is immediately to the north, south, east or west of  $s_i$ . More complex structures are possible in principle, but for simplicity we

restrict ourselves to the binary symmetric scheme given above. Furthermore, we write  $\mathbf{y} = [Y_1, \dots, Y_n]'$  and  $\mathbf{W} = [w_{ij}]$  for  $i, j = 1, \dots, n$ .

Edge points typically raise problems in the statistical analysis of spatial systems. The basic difficulty is that they have fewer neighbours than interior points. For instance, even a simple autoregressive model such as (1) below with  $p = 1$  is not second-order stationary on a finite region without edge corrections (Haining, 1990, p. 82). For this reason we assume that, whenever possible,  $\mathbf{W}$  has been suitably modified to account for edge effects.

In this paper we explore two simple but widely adopted techniques of edge correction. The first one is toroidal correction (Ripley, 1981, p. 152), which wraps a rectangular region onto a torus. Edge points on opposite borders are thus considered to be close, and all sites have the same number of neighbours. In the second instance we apply the asymmetric Neumann correction (Moura and Balram, 1992, p. 338), where the off-region neighbours of a boundary site have the same response value as the site itself. The asymmetric Neumann correction is likely to be preferred in the case of non-lattice data, when the assumption of a toroidal boundary is often not very realistic and may be difficult to implement. We have also tried a modified version of the asymmetric Neumann correction, which might be called "mirror correction", where the off-region neighbours of a boundary site have the same response value as the in-region neighbours of the site. In our applications, however, results from the mirror correction are usually very similar to those obtained under the standard Neumann boundary assumption, and hence will not be reported in detail.

At each location we might have additional (non stochastic) information related to the values of  $p - 1$  spatial covariates. Let  $\mathbf{X}$  denote the corresponding design matrix of dimension  $n \times p$ , allowing also for the mean effect. We consider the simultaneous representation of spatial autoregressive models (SAR):

$$(\mathbf{y} - \mathbf{X}\beta) = (\mathbf{I}_n - \rho\mathbf{W})^{-1}\varepsilon, \quad (1)$$

where  $\beta = [\beta_0, \dots, \beta_{p-1}]'$  is a  $p$ -dimensional parameter vector,  $\mathbf{I}_n$  is the  $n \times n$  identity matrix,  $\rho$  is a measure of spatial interaction between neighbouring sites, and  $\varepsilon = [\varepsilon_1, \dots, \varepsilon_n]'$  is an  $n$ -dimensional vector of disturbances. Errors  $\varepsilon_i$  are defined to be independent and normally distributed with mean 0 and common variance  $\sigma^2$ . For (1) to be meaningful, it is assumed that  $(\mathbf{I}_n - \rho\mathbf{W})^{-1}$  exists.

Estimation of parameters in (1) follows by maximization of the likelihood function

$$l(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-n/2} |\mathbf{I}_n - \rho\mathbf{W}| \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \boldsymbol{\Sigma} (\mathbf{y} - \mathbf{X}\beta)\right\}, \quad (2)$$

where

$$\boldsymbol{\Sigma} = (\mathbf{I}_n - \rho\mathbf{W})' (\mathbf{I}_n - \rho\mathbf{W}),$$

with respect to  $\beta$ ,  $\sigma^2$  and  $\rho$ . This is usually performed in stages, as for a given  $\rho$  the maximum likelihood estimates of  $\beta$  and  $\sigma^2$  are the generalized least squares estimates

$$\hat{\beta} = (\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{y}$$

and

$$\hat{\sigma}^2 = n^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})'\Sigma(\mathbf{y} - \mathbf{X}\hat{\beta}).$$

Given  $\hat{\beta}$  and  $\hat{\sigma}^2$ , the maximum likelihood estimate of  $\rho$ , say  $\hat{\rho}$ , is then obtained by numerical optimization of the profile likelihood derived from (2), subject to the constraint that

$$\hat{\Sigma} = (\mathbf{I}_n - \hat{\rho}\mathbf{W})'(\mathbf{I}_n - \hat{\rho}\mathbf{W})$$

is positive definite. The values of  $\hat{\beta}$ ,  $\hat{\sigma}^2$  and  $\hat{\rho}$  can be easily computed through specialized software, such as module *SpatialStats* of S-plus (Mathsoft, 1996).

In the last two decades the SAR model has been adopted to represent spatial variation in a wide variety of fields. See e.g. Haining (1990); Pace et al. (1998); Griffith and Lane (1999) and Nair et al. (2000) for a number of interesting applications in agriculture, epidemiology, econometrics, socioeconomic analysis and integrated-circuit fabrication.

### 2.2 Spatial outliers

A *spatial outlier* is defined as an observation which is unusual with respect to its neighbouring values (Haining, 1990, p. 214). In the context of the SAR model described above, a common way to assess spatial outlyingness is to compute individual departures from the fitted model. This is accomplished through the vector of standardized residuals

$$\mathbf{e} = \hat{\sigma}^{-1}(\mathbf{I}_n - \hat{\rho}\mathbf{W})(\mathbf{y} - \mathbf{X}\hat{\beta}), \tag{3}$$

which defines the lack of fit statistic  $\mathbf{e}'\mathbf{e}$ . For instance, in Mathsoft (1996, p. 150) it is argued that a plot of standardized residuals against fitted values should provide a valuable check for homogeneity and outliers.

Other definitions of residuals can be useful with correlated data (see e.g., Martin, 1992; Haslett and Hayes, 1998 and Baade and Pettitt, 2000), although our experience with the FS has shown that alternative choices usually provide similar guidance in practice. We prefer standardized over raw residuals as they allow for spatial autocorrelation, which we believe is a sensible property for the purpose of detecting spatial outliers. Furthermore, as in the familiar case of independent observations, the global lack of fit statistic  $\mathbf{e}'\mathbf{e}$  can be decomposed into a sum of individual contributions (i.e. the standardised residuals) which can be identified with the elements of  $\mathbf{y}$ .

A major drawback of standard diagnostic tools is that they are prone to masking and swamping when multiple spatial outliers are present in the data. To show this, we simulated a data set where  $S$  is a  $12 \times 12$  regular grid. Vector  $\mathbf{y}$  was generated according to a SAR model with  $p = 4$ ,  $\rho = 0.2$  and  $\sigma^2 = 0.5$ . Simulation of  $(\mathbf{I}_n - \rho\mathbf{W})^{-1}\varepsilon$  was performed after generation of a normally distributed disturbance vector  $\varepsilon$  and addition of the constant value 20.  $\mathbf{X}$  was also obtained by simulation from a multivariate Normal distribution. The observed  $\mathbf{y}$  was then derived from equation (1) with  $\beta = [20, 5, 4, 3]'$ .

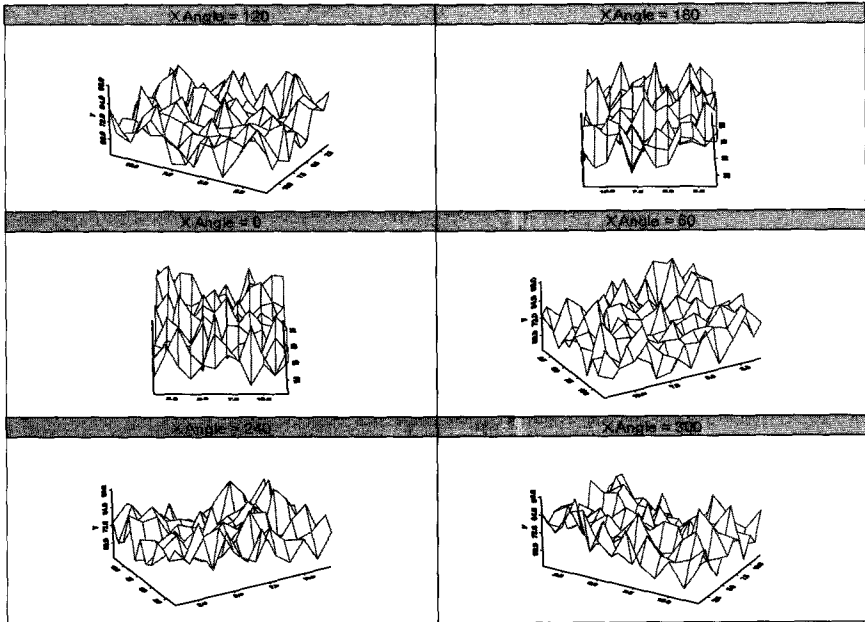
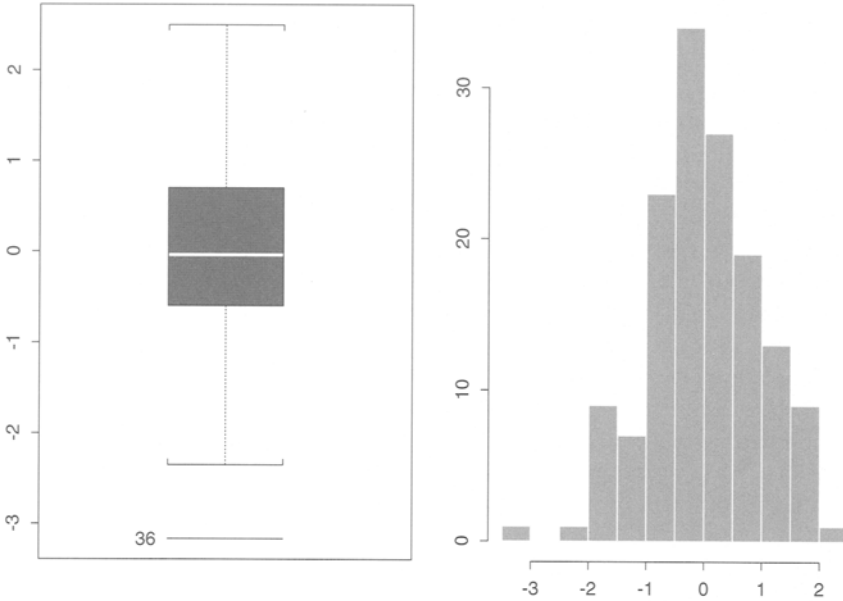


Fig. 1. Simulated example with multiple spatial outliers. Six-panel rotation plot of the data

Response values  $y_i$  were then modified at a  $4 \times 4$  block of sites at the crossings of rows  $1, \dots, 4$  and columns  $1, \dots, 4$ , plus an additional location ( $s_{49}$ , corresponding to the first observation in row 5) on its border. Since contamination was performed after simulation of  $y$ , we can think of the 17 modified values as a cluster of additive spatial outliers. Contamination was not very marked in this example, as it simply amounted to subtracting small constants (ranging from 0.05 to 2.2) from the original readings. Visual inspection of the data from different perspectives (see Fig. 1) does not provide evidence of contamination.

Figure 2 shows the boxplot and an histogram of standardized residuals (3) for this simulated data set with multiple spatial outliers. Masking clearly affects both diagnostic plots and does not allow proper understanding of the features of the data. Indeed, all contaminated values come undetected and only a natural spatial outlier at site  $s_{36}$  is highlighted due to its low standardized residual. The unsatisfactory behaviour of standardized residuals reflects the poor breakdown properties of standard maximum likelihood estimators. An additional problem comes from the fact that the unknown autocorrelation parameter  $\rho$  has been replaced by its sample non-robust estimate  $\hat{\rho}$  in the computation of  $e$ .

Residuals from high-breakdown estimation provide an alternative way to disclose masked features of the data (Rousseeuw and van Zomeren, 1990), although at the cost of reduced efficiency and increased computing time. Indeed, if  $\rho$  were known, estimation of  $\beta$  could follow from standard robust regression analysis. Haining (1990, pp. 384–385) also considered a robust estimator of  $\rho$  based on ordinary least squares. However, this approach is not to be recommended due to the inconsistency of the least squares method under model (1). To our knowledge the



**Fig. 2.** Simulated example with multiple spatial outliers. Boxplot and histogram of standardized residuals

problem of finding a joint robust estimate of  $\rho$  and  $\beta$  has never found a satisfactory solution in practice. The BFS algorithm to be proposed in Sect. 3 specifically addresses this issue.

### 2.3 High leverage sites

Cases for which the regressor values are far from the bulk of the data in the space spanned by the columns of  $\mathbf{X}$  are usually called high leverage points. Detection of such points is an important step of the model building procedure, because they may exert an undue influence on the computed fit. The effect of a high leverage point is to force the fitted model close to the observed value of the response variable. Hence, high leverage points typically have small residuals, even in the absence of masking.

A popular way of measuring individual leverages in standard multiple regression is through the diagonal elements of the hat matrix  $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . Martin (1992) showed how these diagnostics can be extended to cope with dependent observations in the SAR model (1). Specifically, when  $\Sigma \neq \mathbf{I}_n$ , individual leverages can be computed as the diagonal elements of

$$\mathbf{P} = \Sigma\mathbf{X}(\mathbf{X}'\Sigma\mathbf{X})^{-1}\mathbf{X}'\Sigma. \tag{4}$$

Also the diagonal elements of

$$\mathbf{Q} = \Sigma - \mathbf{P}, \tag{5}$$

which are called complementary leverages, can be useful in this context.

Unfortunately, such diagnostics are also prone to masking and swamping effects when multiple high-leverage points are present in the data. To show this, we introduce a new simulated data set, similar to the one analysed in Sect. 2.2. Here  $S$  corresponds to the tiles of a  $20 \times 20$  regular grid, and  $y$  was simulated according to model (1) with  $p = 4$ ,  $\beta = [20, 5, 4, 3]'$  as before,  $\rho = 0.1$  and  $\sigma^2 = 1$ .  $\mathbf{X}$  was also obtained by simulation from a multivariate Normal distribution, but the covariate values at sites  $s_2$ ,  $s_3$  and  $s_{24}$  (in lexicographical order) were now slightly modified in order to increase their leverages.

Figure 3 displays the boxplot and an histogram of the complementary leverages computed from (5) for this data set. Even though the contamination rate is very low (less than 1%), the effects of masking and swamping are clearly apparent, as one sees a plethora of potentially suspect influential observations. In this situation it is difficult to tell what a sensible conclusion would be.

Furthermore, the definition of leverage diagnostics (4) and (5) implies knowledge of  $\rho$ . We have thus to face again the consequences of substituting the non-robust estimate  $\hat{\rho}$  for the unknown autocorrelation parameter.

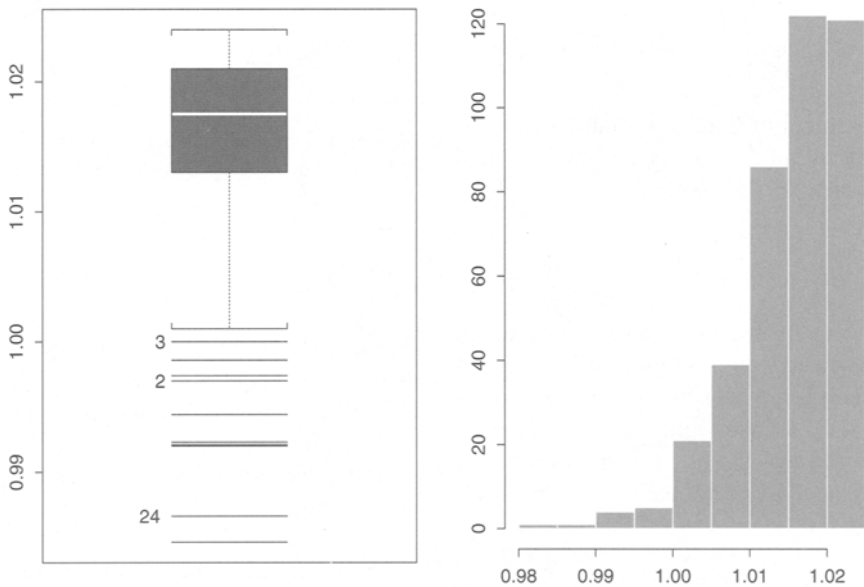


Fig. 3. Simulated example with high leverage points. Boxplot and histogram of complementary leverages

### 3 The block forward search for spatial autoregression

If a simple consistent and efficient estimator of  $\rho$  were known, e.g. a least squares estimator, then it would be possible to compute a robust version of it based on all the data and run a forward search similar to that of Cerioli and Riani (1999).



Unfortunately this solution is not feasible under model (1), since the ordinary least squares estimator of  $\rho$  is not consistent (e.g., Cressie, 1993, p. 407) and generalized least squares estimation requires knowledge of the value of  $\rho$  itself.

We circumvent the difficulty by suggesting a FS algorithm where  $\rho$  is estimated jointly with  $\beta$  at each step of the search. For this purpose, we pick up  $n^* < n$  blocks of contiguous spatial locations within  $S$ , and consider these blocks as the basic elemental sets of our algorithm. Blocks are intended to retain the spatial dependence properties of the whole study region. Hence, they are defined to resemble as closely as possible the shape of  $S$ . Confining attention to subsets of neighbouring locations ensures that spatial relationships are preserved by the BFS algorithm, so that  $\rho$  can be estimated within each block.

In each step of our algorithm parameter estimation is performed through maximum likelihood or a simple approximation to it. Therefore, we are ensured that the estimator of  $\rho$  is consistent at each step of the search, until outliers enter into the fitting subset. Furthermore, for a given  $\rho$ , maximum likelihood provides a more efficient way of estimating  $\beta$  than high-breakdown fitting techniques such as least median of squares.

Let  $A$  be a collection of sites  $s_i \in S$ . Define  $S_A$  to be the subset of  $S$  indexed by locations in  $A$  and denote by  $a$  the cardinality of  $S_A$ . Similarly, take  $\mathbf{y}_A, \mathbf{X}_A, \mathbf{W}_A$  and  $\Sigma_A$  to be the blocks of  $\mathbf{y}, \mathbf{X}, \mathbf{W}$  and  $\Sigma$  corresponding to locations in  $S_A$ . We suppose that also  $\mathbf{W}_A$  is corrected for edge effects through one of the methods described in Sect. 2.1. This adjustment might be particularly important when  $a$  is small.

The exact likelihood function based only on observations in  $S_A$  is described in Martin (1984) and is given by

$$l_A(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-a/2} |(\Sigma^{-1})_A|^{-1/2} \exp \left[ -\frac{1}{2\sigma^2} (\mathbf{y}_A - \mathbf{X}_A\beta)' \{(\Sigma^{-1})_A\}^{-1} (\mathbf{y}_A - \mathbf{X}_A\beta) \right]. \tag{6}$$

In this paper we often work with the simpler approximation

$$l_A^*(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-a/2} |\mathbf{I}_a - \rho\mathbf{W}_A| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y}_A - \mathbf{X}_A\beta)' \Sigma_A (\mathbf{y}_A - \mathbf{X}_A\beta) \right\}, \tag{7}$$

which is defined in terms of  $\Sigma_A$  instead of  $\{(\Sigma^{-1})_A\}^{-1}$ .

Of course, both (6) and (7) correspond to the full likelihood function  $l(\beta, \sigma^2, \rho)$  if  $A = S$ . If  $A \subset S$ ,  $l_A^*(\beta, \sigma^2, \rho)$  yields an approximation which can be evaluated much more quickly than its exact counterpart. Hence, it is particularly appealing for repeated optimization at subsequent steps of the BFS algorithm. In what follows we denote by  $\hat{\beta}_A, \hat{\sigma}_A^2$  and  $\hat{\rho}_A$  the maximizers of either the exact likelihood (6) or its approximation (7), depending on the context.

The main steps of our BFS algorithm for model (1) are summarized as follows.

### 3.1 Defining the blocks

To initialize the FS, we start from a subset of observations which is intended to be outlier free. The initial subset is selected among  $n^*$  blocks  $B_j, j = 1, \dots, n^*$ , of contiguous spatial locations, each of cardinality  $b_j$ . Such blocks are defined to remain as close as possible the same shape as  $S$ . For instance, when  $S$  is a regular square lattice each  $B_j$  is simply a square of  $b_j = c\sqrt{n} \times c\sqrt{n}$  contiguous tiles of  $S$ , for  $c < 1$  a fixed constant.

In the general situation where  $S$  does not have a regular shape, blocks can be obtained as follows. Write  $\mathbf{s}_i$  for the two-dimensional vector of spatial coordinates corresponding to site  $s_i, i = 1, \dots, n$ . For instance, in the case of an administrative unit vector  $\mathbf{s}_i$  may be obtained by simply taking its centroid coordinates. Define  $S^* \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  and  $D(\nu) \equiv (0, \nu_1] \times (0, \nu_2]$ , where  $\nu = \nu_1 \times \nu_2$  is selected in such a way that  $S^* \subseteq D(\nu)$ . Choose a pair of subshape sizes, say  $\phi(\nu_2)$  and  $\phi(\nu_1)$ . Subshapes of  $S^*$  are obtained by dividing the set  $D(\nu)$  into possibly overlapping rectangles each of dimension  $\phi(\nu_1) \times \phi(\nu_2)$ . Then define block  $B_1$  as the subset of spatial locations whose coordinate vector  $\mathbf{s}_i$  lies within the rectangle  $(0, \phi(\nu_1)] \times (0, \phi(\nu_2)]$ . Subsequent blocks are defined in a similar way by simply translating the origin of each rectangle. Since there are only data in  $S^*$ , we only use the rectangles that provide a non-empty intersection with  $S^*$ . Finally, the neighbourhood structure of block  $B_j, j = 1, \dots, n^*$ , is given by  $\mathbf{W}_{B_j}$ .

We suggest to select the subshape sizes  $\phi(\nu_1)$  and  $\phi(\nu_2)$  in such a way that each  $b_j = O(n^{1/2})$ , to ensure a balance between the statistical properties of  $\hat{\rho}_{B_j}$  and the robustness of the method. However, since robustness is our main focus the proportionality constants are likely to be smaller than in other inferential problems for autocorrelated variables where blocks are adopted (e.g. Heagerty and Lumley, 2000; Cerioli, 2002). In fact, blocks are the actual elemental sets of our algorithm and  $n^*$  must be large enough to ensure that there is at least a block without contaminated units. We also avoid the use of overlapping blocks of locations, as this would destroy the diagnostic power of (partially) ordering spatial locations through the BFS algorithm (see Sect. 3.5 below). Some numerical results showing the effect of different choices of  $\phi(\nu_1)$  and  $\phi(\nu_2)$  will be given in Sect. 4.

For simplicity, in what follows we restrict our description of the BFS algorithm to the situation where  $S$  is a regular lattice and  $b_j = b$  for  $j = 1, \dots, n^*$ . This is also the setting of our examples in Sect. 4.2.

### 3.2 Choice of the initial subset

To find the starting subset of the BFS algorithm, we perform an exhaustive search of all possible blocks  $B_j, j = 1, \dots, n^*$ , and choose the one which satisfies a least median of squares criterion. Specifically, let

$$\mathbf{e}_{B_j} = \hat{\sigma}_{B_j}^{-1}(\mathbf{I}_n - \hat{\rho}_{B_j} \mathbf{W})(\mathbf{y} - \mathbf{X}\hat{\beta}_{B_j}),$$

be the  $n \times 1$  vector of standardized residuals computed from the fit to observations in  $B_j$ . This fit can be obtained by maximization of either the exact likelihood function

(6) or its approximate version (7), with  $A = B_j$ . Write  $\mathbf{e}_{B_j}^2$  for the corresponding vector of squared standardized residuals. We select as our initial subset the block, say  $B_*$ , minimizing the function

$$\text{med}(\mathbf{e}_{(2)B_j}^2), \tag{8}$$

where  $\text{med}(\cdot)$  stands for the median of the ordered elements of a vector.

Criterion (8) is similar to the least median of squares method for regression models with independent errors (Rousseeuw, 1984), except that here standardized instead of raw residuals are considered. It should be emphasized that this is the only one step of our algorithm where a high-breakdown criterion is applied. Computation of residuals from a robust fit prevents clusters of spatial outliers from being included in the initial subset. Nevertheless, our experience has shown that this requirement is often less important than expected, since the FS may be able to identify outliers at the outset and replace them with “good” observations.

### 3.3 Progressing in the search

The basic idea of the FS approach is to fit repeatedly the postulated model to subsets of observations of increasing size, selected in such a way that outliers are included only at the end of the procedure. For this purpose, let  $m$  be the number of spatial locations used for fitting model (1) at a step of the BFS. Denote by  $S_{(m)}$  the corresponding subset of  $S$ . At the first step we take  $m = b$  and

$$S_{(m)} = B_*$$

The likelihood at each step is derived from (6) or from its approximation (7) with  $A = S_{(m)}$ . This yields estimates  $\hat{\beta}_{(m)}$ ,  $\hat{\sigma}_{(m)}^2$  and  $\hat{\rho}_{(m)}$ .

$S_{(m)}$  is then updated to  $S_{(m+k(m))}$  by taking the subset of  $m + k_{(m)}$  spatial locations with the smallest absolute standardized residuals in

$$\mathbf{e}_{(m)} = \hat{\sigma}_{(m)}^{-1}(\mathbf{I}_n - \hat{\rho}_{(m)}\mathbf{W})(\mathbf{y} - \mathbf{X}\hat{\beta}_{(m)}), \tag{9}$$

excluding those already in the initial subset. In this search all units but those forming  $B_*$  can thus leave the subset. Of course,  $1 \leq k_{(m)} \leq b$ . For  $m \geq b$  we take either  $k_{(m)} = b$  or  $k_{(m)} = 1$ . In the examples of Sect. 4 both choices give similar results, although the former is often to be preferred as it provides smoother residual trajectories and estimates of  $\rho$ .

We have also implemented an alternative updating scheme where sites from  $B_*$  can be removed from the fitting subset at subsequent steps of the search, a property which was found useful in several applications of the FS run over individual observations (Cerioli and Riani, 1999; Atkinson and Riani, 2000). However, its interpretation might be less intuitive for the BFS described here, since blocks and not individual locations are now the elemental sets. Detailed investigation of this alternative updating scheme will be pursued elsewhere.

3.4 Computing the BFS estimator and relevant diagnostics

The BFS estimator of the multidimensional parameter  $\theta = [\beta', \sigma^2, \rho]'$  is defined as the sequence of its (approximate) maximum likelihood estimators

$$\hat{\theta}_{(m)} = [\hat{\beta}'_{(m)}, \hat{\sigma}^2_{(m)}, \hat{\rho}_{(m)}]'$$

That is,

$$\hat{\theta}_{(FS)} = (\hat{\theta}'_{(b)}, \dots, \hat{\theta}'_{(m)}, \dots, \hat{\theta}'_{(n)})'$$

In addition, a number of diagnostic measures are computed along the BFS. These include standardized residuals (9) for all units, and estimated leverage measures on the diagonal of

$$\hat{\mathbf{P}}_{S(m)} = \hat{\Sigma}_{S(m)} \mathbf{X}_{S(m)} (\mathbf{X}'_{S(m)} \hat{\Sigma}_{S(m)} \mathbf{X}_{S(m)})^{-1} \mathbf{X}'_{S(m)} \hat{\Sigma}_{S(m)} \tag{10}$$

and

$$\hat{\mathbf{Q}}_{S(m)} = \hat{\Sigma}_{S(m)} - \hat{\mathbf{P}}_{S(m)}, \tag{11}$$

for the sites belonging to  $S(m)$ . In both (10) and (11)

$$\hat{\Sigma}_{S(m)} = (\mathbf{I}_m - \hat{\rho}_{(m)} \mathbf{W}_{S(m)})' (\mathbf{I}_m - \hat{\rho}_{(m)} \mathbf{W}_{S(m)}).$$

At each step we also compute the signed square-root likelihood ratio statistic which tests the null hypothesis  $H_0 : \rho = \rho_0$ , for a number of plausible null values  $\rho_0$ . For instance, consider the approximate likelihood (7) and let  $\hat{\beta}_{(m)0}$  and  $\hat{\sigma}^2_{(m)0}$  be the corresponding estimates of  $\beta$  and  $\sigma^2$  computed with  $A = S(m)$  and  $\rho = \rho_0$ . The test statistic is then defined as

$$\begin{aligned} \lambda_{(m)} &= \pm \left[ 2 \log \{ l_{S(m)}^* (\hat{\beta}_{(m)}, \hat{\sigma}^2_{(m)}, \hat{\rho}_{(m)}) \} - 2 \log \{ l_{S(m)}^* (\hat{\beta}_{(m)0}, \hat{\sigma}^2_{(m)0}, \rho_0) \} \right]^{1/2} \\ &= \pm \left\{ \log | \hat{\Sigma}_{S(m)} \Sigma_{S(m)0}^{-1} | - \log (\hat{\sigma}^2_{(m)} / \hat{\sigma}^2_{(m)0}) - \hat{\sigma}_{(m)}^{-2} \mathbf{r}'_{S(m)} \hat{\Sigma}_{S(m)} \mathbf{r}_{S(m)} \right. \\ &\quad \left. + \hat{\sigma}_{(m)0}^{-2} \mathbf{r}'_{S(m)0} \Sigma_{S(m)0} \mathbf{r}_{S(m)0} \right\}^{1/2}, \tag{12} \end{aligned}$$

where the sign is + if  $\hat{\rho}_{(m)} \geq \rho_0$  and - otherwise. In (12),  $\mathbf{r}_{S(m)} = \mathbf{y}_{S(m)} - \mathbf{X}_{S(m)} \hat{\beta}_{(m)}$  is the  $m \times 1$  ordinary residual vector computed from  $\hat{\beta}_{(m)}$  for the units in  $S(m)$ ,  $\mathbf{r}_{S(m)0} = \mathbf{y}_{S(m)} - \mathbf{X}_{S(m)} \hat{\beta}_{(m)0}$  is the corresponding vector under  $H_0$ , and

$$\Sigma_{S(m)0} = (\mathbf{I}_m - \rho_0 \mathbf{W}_{S(m)})' (\mathbf{I}_m - \rho_0 \mathbf{W}_{S(m)}).$$

If  $H_0$  is true  $\lambda_{(m)}$  behaves approximately as a  $N(0, 1)$  random variable. Hence, computation of (12) allows some confirmatory statements about the true value of  $\rho$ . Furthermore, we might expect that the effect of the likelihood approximation (7) will be smaller on  $\lambda_{(m)}$  than on the point estimator  $\hat{\rho}_{(m)}$ .

### 3.5 Diagnostic monitoring and ordering of blocks

A major advantage of the FS approach is to provide the user with a number of informative pictures displaying all the diagnostics computed along the search.

For instance, the anomalous behaviour of spatial outliers is clearly revealed by the individual trajectories in the forward plot of standardized residuals  $e_{(m)}$ , even when standard deletion diagnostics suffer from masking. In a similar fashion, the effect of spatial outliers on the estimate of  $\rho$  can be detected through the forward plots of  $\hat{\rho}_{(m)}$  and  $\lambda_{(m)}$ .

An additional bonus of the FS algorithm is that it is often possible to rank the locations in  $S$  following their entrance step into  $S_{(m)}$ . This gives an ordering of the data according to their degree of agreement with the null model (1), with observations furthest from it joining  $S_{(m)}$  at the last stages of the procedure. However, if the BFS is run with  $k_{(m)} = b$  the ordering is only partial, since all sites belonging to the same block are to be regarded as equivalent. For the same reason, we do not support the definition of overlapping blocks of spatial locations, as in this situation each site would belong to more than one elemental set and ordering would become impossible.

## 4 Applications

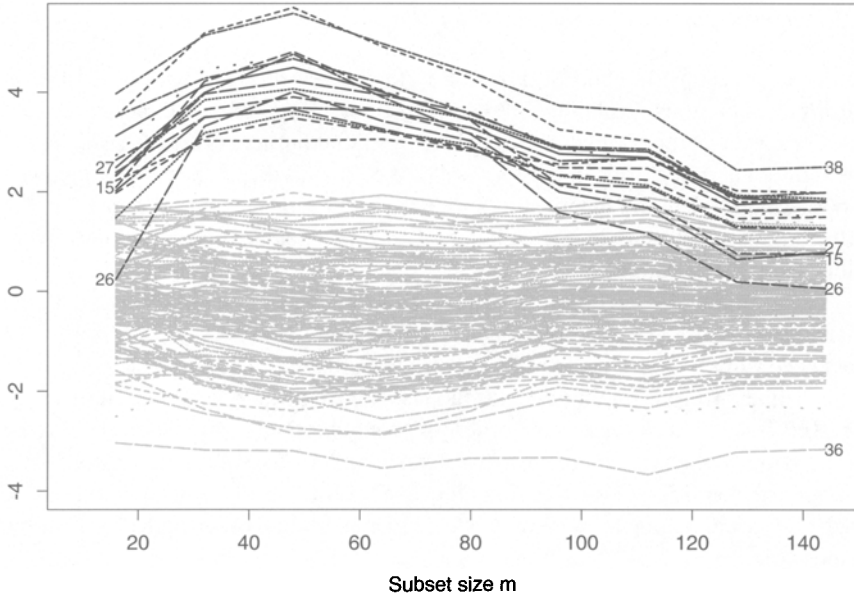
### 4.1 Multiple spatial outliers

We apply our BFS algorithm to the simulated data set with masked spatial outliers described in Sect. 2.2. First, we use toroidal edge corrections and the fast approximation to the likelihood function given in (7). Figure 4 shows the corresponding forward plot of standardized residuals (9) as a function of  $m$ . Here the BFS algorithm is run over non-overlapping square blocks of dimensions  $4 \times 4$  and rule  $k_{(m)} = b$  is adopted for progressing in the search.

The trajectories corresponding to the contaminated cluster clearly stand apart from the others, as these sites have the largest standardized residuals for most of the search. Hence our algorithm is effective in separating outlying locations from the rest of  $S$  even in a situation where contamination is moderate and comes undetected by visual inspection or by traditional diagnostic procedures. In addition, the effect of masking is apparent toward the end of the algorithm, when all residuals have similar magnitude.

It is interesting to see that in this example masking becomes a problem for  $m < n$ , as the modified observations are not particularly different from the bulk of the data and the contaminated corner joins  $S_{(m)}$  at the step prior to the final one. Also residuals from a single robust fit of model (1) might fail in detecting all spatial outliers, as consideration of the least median of squares residuals from the initial subset shows. It is just the *trajectory* of each residual in the forward plot that allows proper understanding of the features of the corresponding location.

To see how our choices affect the results from the search, we also run the BFS algorithm under different settings. Figure 5 provides the forward plot of standardized residuals computed from the approximate likelihood function (7) and the



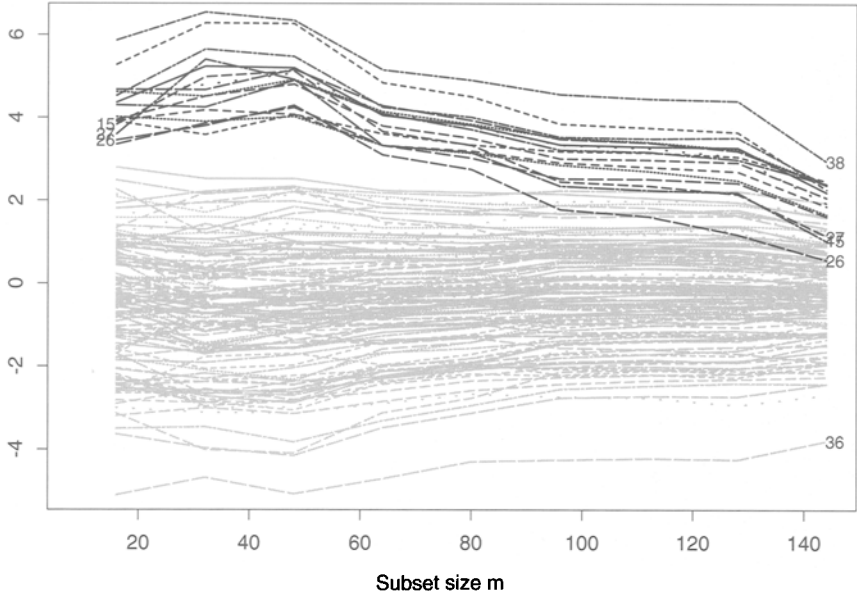
**Fig. 4.** Simulated example with multiple spatial outliers. Trajectories of standardized residuals for individual locations and  $b = 16$ . Curves corresponding to contaminated sites are given in black. *Toroidal edge corrections and approximate maximum likelihood*

asymmetric Neumann edge-correction method introduced in Sect. 2.1. In addition, Fig. 6 is the same plot obtained through the exact likelihood (6) with toroidal edge corrections. In both cases  $b = 16$  as before. It is reassuring to see that the main findings provided by the BFS are the same in all instances, with the contaminated corner clearly standing apart from the remaining trajectories. It is also worth noting that the approximation involved by function (7) is negligible even if the true  $\rho$  is close to its upper limit in this example.

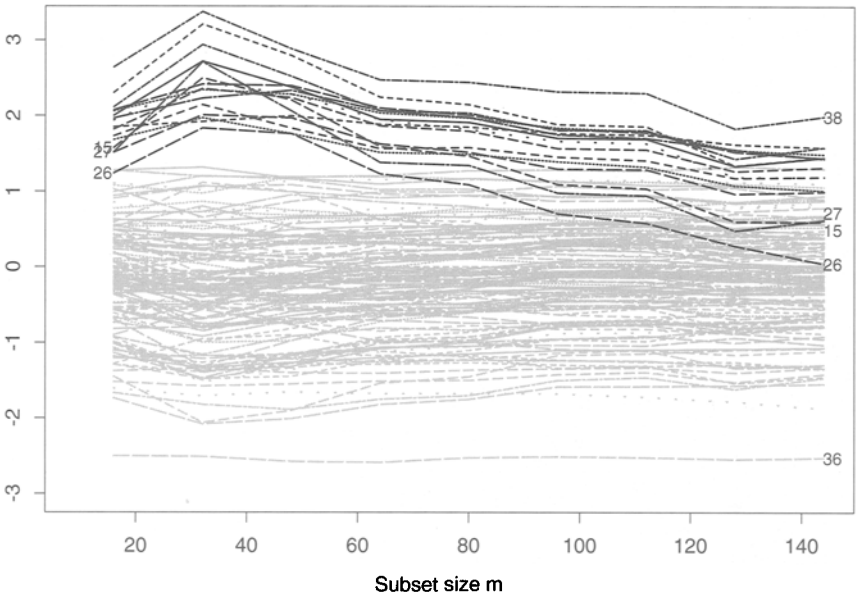
Also the choice of a different block size does not appreciably change the results from the search. For instance, Fig. 7 displays the forward plot of standardized residuals for  $b = 3 \times 3$ , toroidal correction and approximate likelihood. Again, this plot depicts essentially the same information as before, although there is more variability in the first stages of the search, due to the smaller size of  $S_{(m)}$ . The effect of masking also shows up a bit earlier, as spatial outliers are now spread over a larger number of blocks.

#### 4.2 Multiple spatial outliers and estimation of the autocorrelation parameter

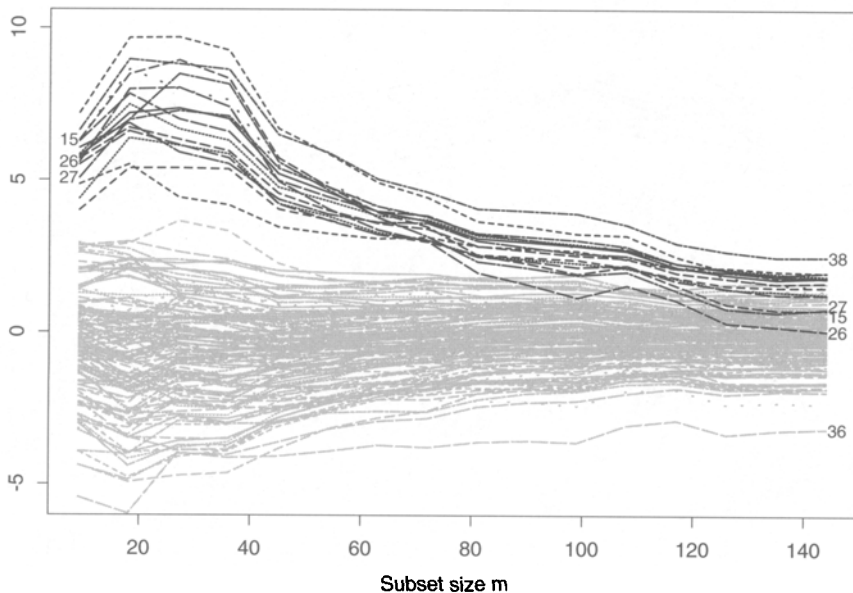
We introduce a new example where multiple spatial outliers have a disproportionate effect on the estimate of  $\rho$ , not only on the residuals from the fitted model. For this purpose, we simulated a data set from model (1) with  $S$  a  $12 \times 12$  grid,  $\beta = [20, 5, 4, 3]'$  as in Sect. 2.2,  $\rho = 0.1$  and  $\sigma^2 = 1$ . Then we modified  $y_i$  at a block of 16 sites located at the crossings of rows  $1, \dots, 4$  and columns  $1, \dots, 4$ , by subtracting



**Fig. 5.** Simulated example with multiple spatial outliers. Trajectories of standardized residuals for individual locations and  $b = 16$ . Curves corresponding to contaminated sites are given in black. *Asymmetric Neumann edge corrections and approximate maximum likelihood*



**Fig. 6.** Simulated example with multiple spatial outliers. Trajectories of standardized residuals for individual locations and  $b = 16$ . Curves corresponding to contaminated sites are given in black. *Toroidal edge corrections and exact maximum likelihood*



**Fig. 7.** Simulated example with multiple spatial outliers. Trajectories of standardized residuals for individual locations and *reduced block size* ( $b = 9$ ). Curves corresponding to contaminated sites are given in black. *Toroidal edge corrections and approximate maximum likelihood*

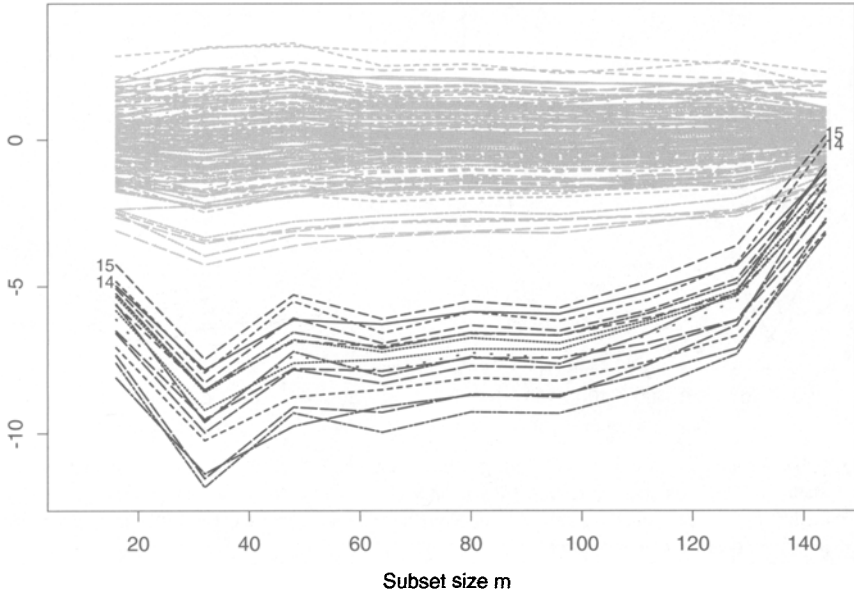
8 from all of them. Fixed contamination increases the similarity of response values in the outlying corner and thus has a larger influence on estimation of  $\rho$ .

Although in this example contamination on the response variable is marked and perhaps it might be detected also by other methods, the BFS still provides important additional information through the display of residual trajectories and the monitoring of diagnostic quantities. Furthermore, it is the only robust method that is able to highlight the inferential effect of each observation on the estimate of the spatial autocorrelation parameter of a SAR model.

The BFS is applied here with  $b = 4 \times 4$  and  $k_{(m)} = b$ , although similar results are obtained with different block sizes. First, we adopt toroidal correction and the approximation (7) to the likelihood function of blocks. As in our example of Sect. 4.1, contaminated locations clearly stand out in the forward plot of standardized residuals (Fig. 8). Masking is still present in the final step, when  $S_{(m)} = S$ , and particularly so for the observations at sites  $s_{14}$  and  $s_{15}$ .

The left panel of Fig. 9 shows the forward plot of  $\hat{\rho}_{(m)}$ , again as a function of  $m$ . The right panel gives the forward plot of the signed square-root likelihood ratio statistic (12), which tests  $H_0 : \rho = \rho_0$  for a number of values  $\rho_0 > 0$ , together with asymptotic 99% confidence bands. The effect of including the contaminated subset is paramount in the final step of both displays, raising the estimate of the autocorrelation parameter from  $\hat{\rho} = 0.103$  based on  $S_{(128)}$  to  $\hat{\rho} = 0.223$  based on all the data. Multiple outliers grossly mislead confirmative analysis based on the likelihood ratio statistic, with the true value  $\rho = 0.1$  being wrongly rejected only





**Fig. 8.** Second simulated example with multiple spatial outliers. Trajectories of standardized residuals for individual locations. Curves corresponding to contaminated sites are given in black. *Toroidal edge corrections and approximate maximum likelihood*

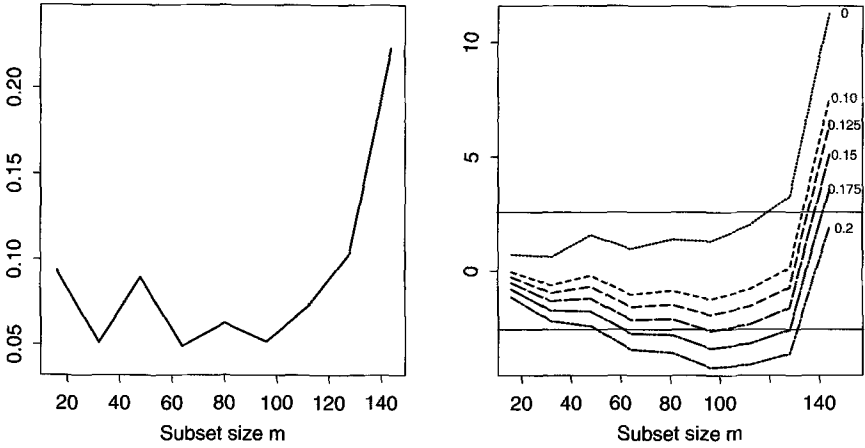
after their inclusion. On the other hand, it is seen that the apparently more plausible value  $\rho = 0.2$  lies well outside the 99% confidence bands for most of the search.

Again, our findings on  $\rho$  are not affected by the specific edge correction method, nor by use of the approximate likelihood function (7) instead of its exact counterpart. Figure 10 displays the forward plots of  $\hat{\rho}_{(m)}$  and the signed square-root likelihood ratio statistic  $\lambda_{(m)}$  computed under asymmetric Neumann correction and approximate likelihood. Similarly, Fig. 11 displays the same graphs for toroidal correction and exact likelihood. It is clearly seen that these pictures provide essentially the same information as Fig. 9. The only difference is the more accurate estimate of  $\rho$  obtained through the exact likelihood function (6) when  $m \approx n/2$ .

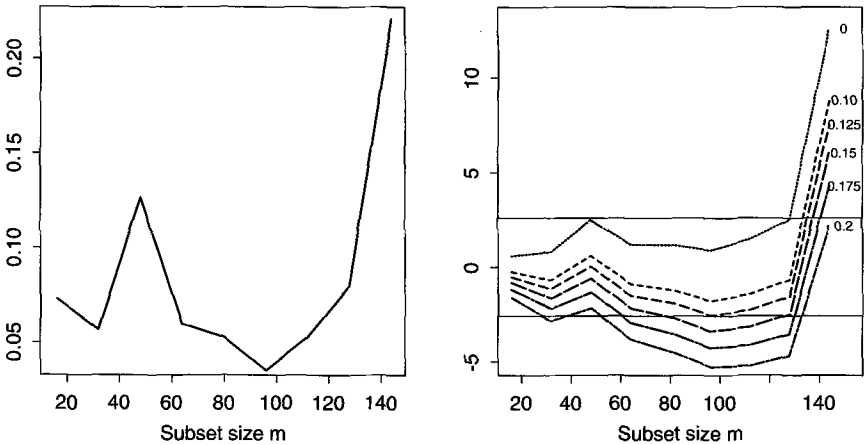
We conclude this example by stressing the value of the BFS algorithm for the purpose of detecting the influence of multiple outliers on estimation of  $\rho$ . Furthermore, outliers enter towards the end of the algorithm in a step that can often be easily identified by our diagnostic plots (see the marked elbow in the curve of  $\hat{\rho}_{(m)}$ ). Quantities computed before their inclusion can thus be regarded as robust against them.

### 4.3 High leverage sites

We now apply the BFS algorithm to the simulated data set with masked high-leverage points described in Sect. 2.3. The procedure is run over non-overlapping square blocks of dimensions  $4 \times 4$ , so that  $b = n/9$ . Rule  $k_{(m)} = 1$  is adopted here



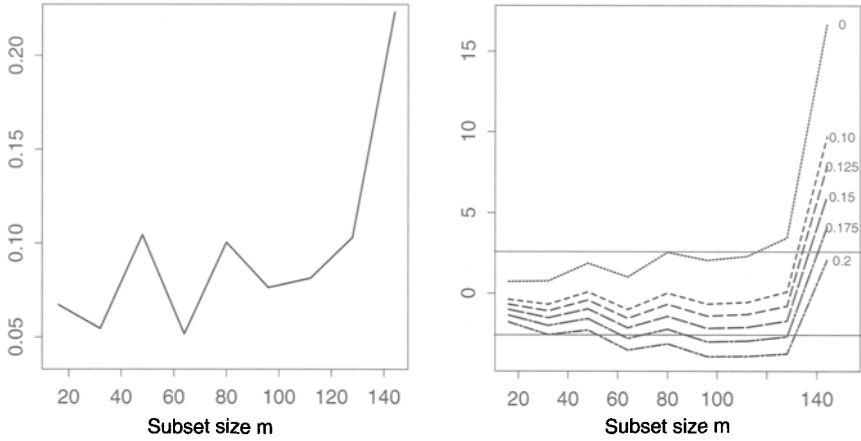
**Fig. 9.** Second simulated example with multiple spatial outliers. Estimates of  $\rho$  (left) and signed square-root likelihood ratio statistics for testing  $H_0 : \rho = \rho_0$ , for a number of values  $\rho_0$ , (right) at each step of the BFS. In the right panel the 99% confidence bands are also reported. *Toroidal edge corrections and approximate maximum likelihood*



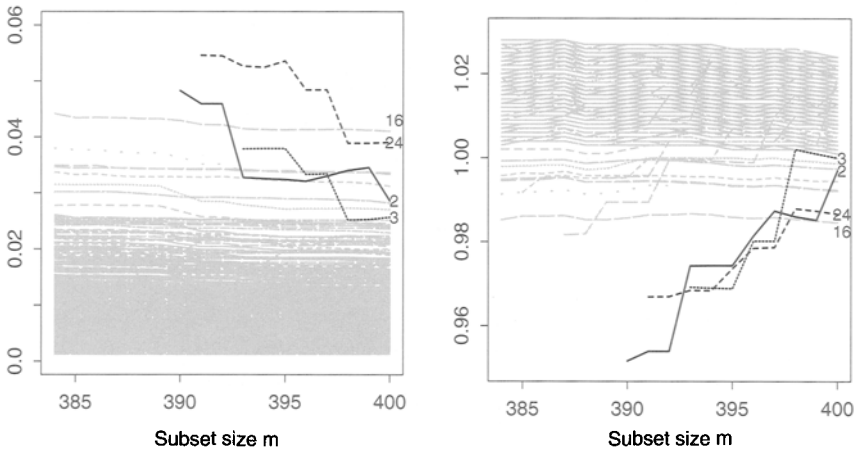
**Fig. 10.** Second simulated example with multiple spatial outliers. Estimates of  $\rho$  (left) and signed square-root likelihood ratio statistics for testing  $H_0 : \rho = \rho_0$  (right) at each step of the BFS. *Asymmetric Neumann edge corrections and approximate maximum likelihood*

for progressing in the search. This choice is motivated by the need to obtain detailed information about the behaviour of individual leverage values with respect to other locations within the same block. With large subset sizes, repeated evaluation of the exact likelihood function (6) becomes computationally demanding. For simplicity, we thus restrict ourselves here to the fast approximation given in (7) with toroidal edge corrections.

Figure 12 displays the resulting forward plots of estimated leverages (left) and complementary leverages (right), as a function of the size of  $S_{(m)}$ , in the last 20 steps of the search. Each site is monitored following its inclusion into  $S_{(m)}$ . The



**Fig. 11.** Second simulated example with multiple spatial outliers. Estimates of  $\rho$  (left) and signed square-root likelihood ratio statistics for testing  $H_0 : \rho = \rho_0$  (right) at each step of the BFS. *Toroidal corrections and exact maximum likelihood*



**Fig. 12.** Simulated example with high leverage points. Estimated leverages (left) and complementary leverages (right) in the last 20 steps of the search. Curves corresponding to contaminated high-leverage sites are given in black. *Toroidal edge corrections and approximate maximum likelihood*

curves corresponding to the three faked high-leverage points are highlighted in the plot. The effect of masking appears toward the end of the algorithm, when these sites have leverage measures comparable to those of other locations.

Masking may be a bit surprising for site  $s_{24}$ , which is not a neighbour of  $s_2$  and  $s_3$  according to our definition of  $\mathbf{W}$ . Nevertheless, monitoring leverage diagnostics at subsequent steps of the BFS clearly shows a different behaviour for all the three sites that were contaminated in the space of explanatory variables. Although both panels of Fig. 12 convey essentially the same information, the picture on the left shows smoother trajectories for uncontaminated sites.

It is worth noting that the estimated leverage for site  $s_3$  is always less extreme than the estimated leverage for  $s_{16}$ , a natural high-leverage site, and becomes almost unnoticeable in the last few steps of the search. It is thus the *shape* of the corresponding trajectory that clearly leads to proper recognition of the peculiarity of this point in the (generalized) space spanned by the columns of  $\mathbf{X}_{S(m)}$ . On the contrary, the trajectory of  $s_{16}$  is quite similar to that of the bulk of the other locations, suggesting that the remoteness of this site in the regressor variable space does not depend on the inclusion of the units which enter in the last steps of the BFS algorithm.

#### 4.4 Wheat yield data

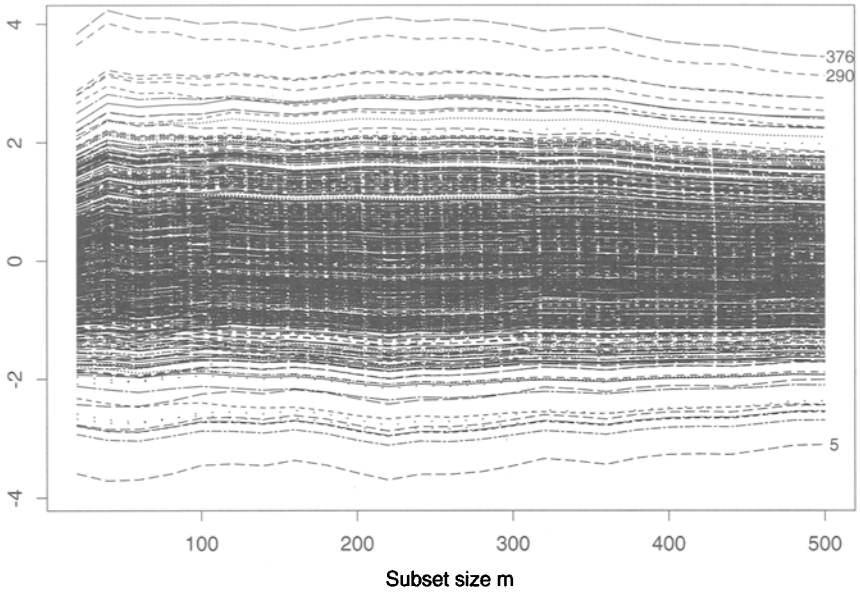
As a final example, we apply the BFS algorithm to a real data set. For this purpose we consider the wheat yield originally reported by Mercer and Hall on a rectangular grid of dimension  $20 \times 25$ : see e.g. Cressie (1993, Sect. 7.1) for a detailed description of the data. Although there is some evidence that a one-parameter spatial autocorrelation model such as (1) might not be appropriate in this application, the Mercer and Hall data have been widely used for the purpose of describing spatial a utoregression. This application is thus intended to provide a comparison of results from our BFS algorithm with those obtained by more standard techniques. Furthermore, Riani and Cerioli (2002) showed through the FS for kriging that the wheat-yield data set does not seem to be affected by masking, but only contains a few isolated spatial outliers. Hence it is useful to check that the BFS does not introduce spurious information with relatively “well behaved” spatial data.

In this example there are no explanatory variables, so that model (1) is purely spatial,  $p = 1$  and the leverage measures discussed in Sects. 2.3 and 4.3 are not relevant.

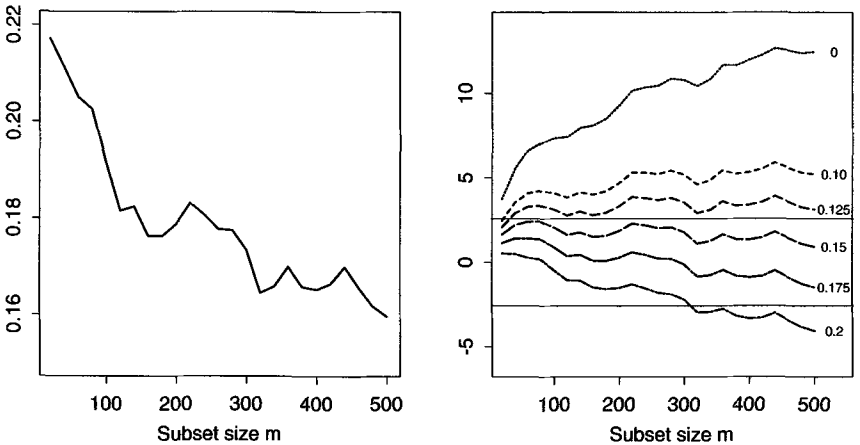
The BFS is run over non-overlapping rectangular tiles of contiguous spatial locations. In what follows we take blocks of dimensions  $b = 4 \times 5$  and update  $S(m)$  according to rule  $k(m) = b$ . For computational simplicity, we again restrict ourselves to the fast approximation to the likelihood function given in (7) with toroidal edge corrections.

Figure 13 shows the forward plot of standardized residuals (9) as a function of the size of  $S(m)$ . Individual trajectories have a regular and homogeneous behaviour, with only three possible outliers. Since these locations are scattered within  $S$ , there is no masking toward the end of the algorithm. The number of curves in the top half-plane of the plot is approximately equal to  $n/2$ , so that there is no evidence of a spatial trend. The information gained by the BFS algorithm for the SAR model is similar to that obtained by Riani and Cerioli (2002) within a geostatistical framework, although a different distance scheme is adopted here and the spatial autocorrelation parameter  $\rho$  is estimated at each step of the search. On the contrary, exploratory data analysis performed by Cressie (1993) through more classical geostatistical methods led to questionable detrending and introduced spurious outliers.

Figure 14 shows the forward plots of  $\hat{\rho}(m)$  and  $\lambda(m)$ , for a number of null values  $\rho_0 > 0$ . From both panels we see the influence of each block of spatial locations



**Fig. 13.** Wheat yield data. Trajectories of standardized residuals for individual locations. Toroidal edge corrections and approximate maximum likelihood



**Fig. 14.** Wheat yield data. Estimates of  $\rho$  (left) and signed square-root likelihood ratio statistics for testing  $H_0 : \rho = \rho_0$ , together with 99% confidence bands, (right) at each step of the BFS. Toroidal edge corrections and approximate maximum likelihood

on estimation of  $\rho$ . The final estimate  $\hat{\rho}_{(n)} = 0.16$  corresponds to that obtained through the standard non-robust fitting technique. As the search stabilizes, values of  $\rho$  belonging to the interval  $(0.15; 0.175)$  become increasingly plausible. Hence in this application the blocks that are included in the last steps of the search have only a minor effect on the estimated autocorrelation parameter. This is a sensible behaviour in the absence of clusters of masked spatial outliers.

### 5 Final remarks

In this paper we have shown how the Forward Search algorithm for regression models with independent errors (Atkinson and Riani, 2000) can be extended to spatial autoregressive models with simultaneous specification. The main contribution of our approach is that it allows joint robust estimation of both trend and autocorrelation parameters. This result follows from a novel definition of the elemental sets of the Forward Search, which is run over blocks of contiguous spatial locations.

We have shown through a number of examples that our algorithm is a valuable tool for the purpose of detecting multiple outliers and high leverage points which would be masked by standard diagnostic methods. This is achieved through joint consideration of a number of revealing plots. In contrast to other robust fitting techniques, which are not easily extended to spatial models, it has the additional advantage of relying on efficient and easy-to-implement estimation schemes, such as maximum likelihood. Furthermore, it fulfills the difficult task of providing some useful robust diagnostics for the estimated autocorrelation parameter that would not be available otherwise. We have also seen that this power is not diminished by adopting a fast approximation to the likelihood function of each block, nor by choosing a number of alternative options in the edge correction procedure and in the search algorithm.

However, we have not answered all relevant questions. An important extension of our BFS algorithm is to the fitting of SAR models with directional and/or second-order effects, as well as to the conditional autoregressive (CAR) model

$$(\mathbf{y} - \mathbf{X}\beta) = (\mathbf{I}_n - \rho\mathbf{W})^{-1/2}\varepsilon. \tag{13}$$

Indeed, in the case of a CAR model an interesting research problem arises. On the one hand, the BFS algorithm could be simply modified by taking the likelihood function for a CAR model

$$l^{CAR}(\beta, \sigma^2, \rho) = (2\pi\sigma^2)^{-n/2}|\Xi|^{1/2}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'\Xi(\mathbf{y} - \mathbf{X}\beta)\right\}, \tag{14}$$

where

$$\Xi = \mathbf{I}_n - \rho\mathbf{W}$$

with  $\mathbf{W}$  symmetric, as required by the CAR assumption. The estimates  $\hat{\beta}_{(m)}$ ,  $\hat{\sigma}_{(m)}^2$  and  $\hat{\rho}_{(m)}$  computed at subsequent steps of the search could then be obtained from the CAR likelihood (14) based only on observations in  $S_{(m)}$ .

On the other hand, ordinary least squares yields a consistent estimate of  $\rho$  under model (13). Hence it is possible, in principle at least, to compute a robust estimate of the spatial autocorrelation parameter based on all the data and run a FS conditional on it, as we did for kriging (Cerioli and Riani, 1999). CAR models will thus provide the ideal ground where the results from the BFS approach of this paper and those from the conditional FS for kriging can be compared in terms of their statistical and practical efficiency.

Robust estimation of the Box-Cox transformation parameter within the framework of model (1) is an additional point which seems a particularly promising

consequence of the forward approach of this paper (Riani and Atkinson, 2000). Indeed, for the purpose of finding an adequate data transformation, either ignoring the effect of spatial autocorrelation or trying to eliminate it have been customary practices in the past (e.g. Haining, 1990, p. 227). Our BFS algorithm, on the contrary, has the great potential of allowing joint (maximum likelihood) estimation of both autocorrelation and transformation parameters. How to solve this and the other aforementioned problems is currently under investigation.

*Acknowledgements.* The authors are grateful to Anthony C. Atkinson and to two anonymous reviewers for useful comments on an earlier draft of the paper. This research was partially supported by grants from Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR).

## References

- Atkinson AC (1994) Fast very robust methods for the detection of multiple outliers. *Journal of the American Statistical Association* **89**, 1329–1339
- Atkinson AC, Riani M (2000) *Robust Diagnostic Regression Analysis*. Springer, New York
- Atkinson AC, Riani M, Cerioli A (2003) *Robust Diagnostic Multivariate Data Analysis*. Springer-Verlag, New York (in preparation)
- Baade A, Pettitt AN (2000) Multiple and conditional deletion diagnostics for general linear models. *Communications in Statistics – Theory and Methods* **29**, 1899–1910
- Cerioli A (2002) Testing mutual independence between two discrete-valued spatial processes: a correction to Pearson chi-squared. *Biometrics* **58**, 888–897
- Cerioli A, Riani M (1999) The ordering of spatial data and the detection of multiple outliers. *Journal of Computational and Graphical Statistics* **8**, 239–258
- Cressie NAC (1993) *Statistics for spatial data*. Wiley, New York
- Griffith DA, Layne LJ (1999) *A casebook for spatial statistical data analysis*. Oxford University Press, New York
- Hadi AS (1992) Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society, Series B* **44**, 761–771
- Hadi AS, Simonoff JS (1993) Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association* **88**, 1264–1076
- Haining R (1990) *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge
- Haslett J, Hayes K (1998) Residuals for the linear model with general covariance structure. *Journal of the Royal Statistical Society, Series B* **60**, 201–215
- Heagerty PJ, Lumley T (2000) Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association* **95**, 197–211
- Martin RJ (1984) Exact maximum likelihood for incomplete data from a correlated Gaussian process. *Communications in Statistics – Theory and Methods* **13**, 1275–1288
- Martin RJ (1992) Leverage, influence and residuals in regression models when observations are correlated. *Communications in Statistics – Theory and Methods* **21**, 1183–1212
- Mathsoft (1996) *S+SpatialStats. User's Manual*. MathSoft Inc. Seattle
- Moura JMF, Balam N (1992) Recursive structure of noncausal Gauss-Markov random fields. *IEEE Transactions on Information Theory* **38**, 334–354
- Nair V, Hansen M, SHIJ (2000) Statistics in advanced manufacturing. *Journal of the American Statistical Association* **95**, 1002–5005
- Pace RK, Barry R, Sirmans CF (1998) Spatial statistics and real estate. *Journal of Real Estate Finance and Economics* **17**, 5–13
- Politis DN, Romano JP, Wolf M (1999) *Subsampling*. Springer, New York
- Riani M, Atkinson AC (2000) Robust diagnostic data analysis: Transformations in regression. *Technometrics* **44**, 384–391

- Riani M, Cerioli A (2002) La forward search per il controllo di qualità di dati autocorrelati. Lauro NC e Scepi G (eds.): *Analisi Multivariata per la Qualità Totale Metodologia, aspetti computazionali ed applicazioni* (in Italian) Franco Angeli, Milano, (in press)
- Ripley BD (1981) *Spatial Statistics*. Wiley, New York
- Rousseeuw PJ (1984) Least median of squares regression. *Journal of the American Statistical Association* **79**, 871–880
- Rousseeuw PJ, van Zomeren BC (1990) Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* **85**, 633–639
- Sherman M (1996) Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society, Series B* **58**, 509–523