GENERAL PAPERS

Spatial linear models with autocorrelated error structure

S. RICHARDSON,[1] C. GUIHENNEUC[2] & V. LASSERRE[2]

[1] *INSERM U.170, 16 Avenue Paul Vaillant-Couturier, 94807 Villejuif, France*
[2] *Laboratoire de Statistique Médicale, CNRS-URA 13-23, Université de Paris V, 45 rue des Saints-Pères, 75006 Paris, France*

Abstract. This paper considers spatial regression problems for irregularly spaced data points. The choice of the parametric spatial model for the residuals and its influence on the testing of the regression coefficients is discussed in a maximum likelihood framework. An iterative estimation procedure based on estimated generalized least squares is also defined and some of its drawbacks are outlined. The various methods of analysis are compared in an example concerned with male lung cancer rates and industry in France.

1 Introduction

Maximum likelihood estimation in spatial regression problems is currently receiving a lot of attention (Warnes & Ripley, 1987; Mardia & Watkins, 1989; Mardia, 1990). The framework is the following: we have a Gaussian process $\{Y(i), i \in A\}$ defined on a domain A containing N locations. We have a set of non-random regressors $X = \{X_1, \ldots, X_p\}$ defined throughout A and we suppose that:

$$Y = X\beta + U, \tag{1}$$

where β is a $p \times 1$ vector of parameters and U follows a multinormal distribution $N(0, \Sigma)$.

This framework is quite general and could include in particular trend surface analysis where X is a vector of spatial coordinates. Spatial regression problems arise in many domains of science such as the earth sciences (Malin & Hide, 1982; Agterberg, 1984), geography (Cliff & Ord, 1981; Bivand, 1984; Haining, 1990), agricultural field trials (Martin, 1990), econometrics (Anselin, 1988) or epidemiology (Cook & Pocock, 1983).

The log likelihood can be written (up to an additive constant) as:

$$L(\beta, Y) = -\tfrac{1}{2}\log|\Sigma| - \tfrac{1}{2}(Y - X\beta)'\Sigma^{-1}(Y - X\beta).$$

For fixed Σ, maximizing over β gives $\hat{\beta} = (X'\Sigma^{-1}X)^{-1}\, X'\Sigma^{-1}Y$, which has the form of the generalized least squares (GLS) estimate of β.

Since the matrix Σ is only rarely known, tests of the regression coefficient β are done conditional on an estimated structure for Σ. The matrix Σ might be modelled as $\Sigma = \Sigma(\theta)$ in terms of a q-dimensional vector of parameters θ. In this case the maximum likelihood estimation of θ is the value that globally maximizes the profile likelihood:

$$L_{\mathrm{p}}(\theta, Y) = -\tfrac{1}{2}\log|\Sigma| + \tfrac{1}{2}(Y - X\hat{\beta})'\Sigma^{-1}(Y - X\hat{\beta}). \tag{2}$$

Maximum likelihood estimation thus involves, in the first place, the choice of a parametric model for Σ and, in a second stage, the numerical maximization of the profile likelihood L_{p}. Following the spatial models chosen for Σ, the profile likelihood L_{p} might be multimodal and detection of the maximum requires careful numerical analysis as emphasized by Ripley (1988). Alternatively Σ might be estimated directly and the regression coefficients evaluated by an iterative estimated generalized least squares (EGLS) procedure similar to the one proposed by Cochrane & Orcutt (1949) for time series. Other types of spatial covariance estimators have been discussed by Zimmerman & Zimmerman (1991) in the different context of prediction and kriging.

The problem of choice of spatial model for the residuals and of its influence on the testing of the regression coefficients has received little attention, particularly when the data are non-lattice. Haining (1987) compares characterizations of Σ in connection with trend surface estimation in the case of regularly spaced aerial survey data. For agricultural field trials, Martin (1990) discusses model selection by extending techniques from time series modelling among a class of spatial processes with row and column structure. We have chosen to investigate the influence of the choice of models on the estimation and testing of β; first in the maximum likelihood framework by considering, for the same examples, different parametric models for U or Σ and secondly by discussing the results given by EGLS with direct estimation of Σ.

A modified test of the partial correlation between Y and X_k conditional on $\{X_l, l \neq k\}$ has been proposed by Richardson (1990). It allows for the presence of a spatial structure in Y and X without requiring the use of parametric modelling. Modified tests of partial correlation will be calculated on the same examples and compared with the tests of the regression coefficients given by the different models.

The examples concern the geographical distribution of male lung cancer in France and its link with industrial risk factors. Apart from smoking, this cancer site has been often linked with occupational exposure in individuals (Simonato *et al.*, 1988) and it is thus interesting to see how this association is reflected geographically.

A common feature of spatially located mortality data is that it often exhibits some degree of positive spatial autocorrelation. This autocorrelation has a different interpretation depending on the type of disease investigated. The spatial structure of infectious diseases reflects the direct interaction of cases in neighbouring areas. In contrast, for chronic diseases, this autocorrelation can be thought of as arising indirectly through the influence on the disease of many covariates which may vary at different spatial scales, typically larger than those under consideration. Some of these covariates, if they are identified and measurable, can be introduced in a regression model as will be done in our examples. Nevertheless the residual variations might still be spatially autocorrelated. In a multifactorial chronic disease like lung cancer, this residual autocorrelation gives an indication that other risk factors, which possess a spatial structure, are potentially influencing the geographical variations of lung cancer mortality. For example, lung cancer has been associated with air pollution or exposure to radon but, depending on the geographical scale of the study, these environmental risk factors could be very difficult to quantify meaningfully. Hence, the modelling of the spatial structure of the residuals will in effect summarize all the unidentified or unmeasurable confounding risk factors and allow a correct statistical inference on those which are included in the regression model.

2 Models and methods

2.1 Spatial characterization of the variance–covariance matrix of the residuals

2.1.1 Parametric models for U. Spatial autoregressive models for U have been proposed by analogy to the commonly used time series method of ARMA modelling of residuals errors in regression (Ord, 1975). For irregularly spaced data points, a known $N \times N$ matrix of weights W representing contiguity between the spatial locations is usually defined and then simultaneous (SAR) or conditional (CAR) models are defined.

A SAR Gaussian model (Whittle 1954) for $\{U_i, 1 \leqslant i \leqslant N\}$ is defined by N simultaneous equations:

$$U_i = b \sum_j W_{ij} U_j + \varepsilon_i, \qquad 1 \leqslant i \leqslant N, \tag{3}$$

where the parameter b represents an overall spatial dependence and the $\{\varepsilon_i, 1 \leqslant i \leqslant N\}$ are independent Gaussian variables with zero mean and variance σ^2. The covariance matrix Σ

is then equal to $\sigma^2(I-bW)^{-1}(I-bW')^{-1}$. We shall consider here the case where the matrix W has been normalized with row sums equal to 1. Note that W is no longer symmetric.

CAR Gaussian models (Besag, 1974) specify the conditional distribution of U_i given all other $U_j, j \neq i$ as Gaussian with:

$$\begin{cases} E[U_i|U_j, j\neq i] = c\sum_j W_{ij}U_j \\ V[U_i|U_j, j\neq i] = \sigma^2. \end{cases} \tag{4}$$

In this case W has to be symmetric and Σ is equal to $\sigma^2(I-cW)^{-1}$. For this model, we took W to be the original 0–1 contiguity matrix. Examples of the fitting of SAR or CAR models are given in Cliff &Ord (1981), Doreian (1981), Upton & Fingleton (1985), Haining (1987), Cressie & Chan (1989) and Mollié & Richardson (1991).

2.1.2 Parametric models for Σ. As an alternative to specifying models for U, the spatial characteristics of the variance covariance matrix Σ can be directly modelled with the help of a small number of parameters. Ripley (1981) and Mardia & Watkins (1989) discuss classes of isotropic spatial covariance functions which ensure that Σ is always non-negative definite. One family of functions, often referred to as the disc model, defines the covariance between the two points (i,j) with distance d_{ij} between them as proportional to the intersection area of two discs of common radius centred on those points.

When d_{ij} is the distance between the sites i and j, the (i,j) element of Σ is then given by $\sigma^2 f_a\,(d_{ij})$ with:

$$f_a(r) = \frac{2}{\pi}\left[\cos^{-1}\left(\frac{r}{2a}\right) - \frac{r}{2a}\left(1-\frac{r^2}{4a^2}\right)^{1/2}\right], \qquad r \leqslant 2a \tag{5}$$

and $f_a(r)$ equal zero elsewhere.

This covariance function, which depends on only one parameter a, exhibits a fairly linear decrease with increasing distance (Ripley, 1981) with the value for $f'_a(0)$, the slope of the tangent at zero, equal to $-2(a\pi)^{-1}$. By considering the intersection of two spheres, a three-dimensional analogue of the disc model can be similarly defined.

Another family of functions, first introduced by Whittle in 1954, defines the covariance between two points (i,j) with the help of a two parameter family of functions as $\sigma^2\, g_{v,\delta}\,(d_{ij})$ with:

$$g_{v,\delta}(r) = [2^{v-1}\Gamma(v)]^{-1}\,(\delta r)^v K_v(\delta r), \quad v>0, \quad \delta>0, \tag{6}$$

where K_v are the modified Bessel functions. The parameter v is a shape parameter whilst δ is a spatial parameter conditioning how far non zero autocorrelations will stretch. Setting $v=\frac{1}{2}$ gives an exponential correlation function depending only on one parameter. A more general form of exponential decrease depending on two coefficients can also be defined and was used by Cook and Pocock (1983). The (i,j) element of Σ is given by:

$$\sigma^2\gamma\exp(-\lambda d_{ij}). \tag{7}$$

Including the possibility of anisotropy, Vecchia (1988) considered a general form for a rational spectral density function of a two-dimensional process, whose covariance can be expressed in terms of derivatives of the Bessel function $K_0(\delta r)$. This family of covariances overlaps partially with those introduced by Whittle. In our comparative study of different parametric models for Σ, we shall restrict ourselves to the five classes of parametric models defined in (3) to (7).

Other functional forms for Σ that do not necessarily ensure that Σ is non-negative definite have also been tried. This is the case for the quadratic distance function proposed by Agterberg (1984) and used by Haining (1987) for modelling the autocorrelation in trend analysis.

2.1.3 *Direct estimation of* Σ. Since there are no repeated measures at the same locations, some assumptions have to be made to be able to estimate Σ directly in a framework which is not model dependent. In the case of an irregular domain, a common assumption is that the set of pairs $\{(i,j)\in A\times A\}$ can be divided into strata S_k, $0\leqslant k\leqslant K$, $S_0=\{(i,i),\ i\in A\}$, of cardinal N_k so that the covariance between U_i and U_j is constant in each strata, i.e. $\mathrm{cov}(U_i, U_j)=C(k)$ for $(i,j)\in S_k$. If isotropy is assumed then $C(k)$ can be indexed by a discrete distance function and this is the assumption that we shall make henceforth for simplicity, though non-isotropic covariances can easily be accommodated.

Estimates of the covariances $C(k)$ need to be considered now in order that the matrix Σ_D with elements $\{\sigma_{ij}, (i,j)\in A\times A\}$ defined by $\sigma_{ij}=C(k)$ if $(i,j)\in S_k$, can be estimated. They will be based on the vector of residuals: $\hat{U}=Y-X\hat{\beta}$, where $\hat{\beta}$ is the ordinary least squares (OLS) estimate of β at a first step and subsequently an iterative EGLS estimator. In line with the assumptions made above, these estimates will involve, the sums

$$\sum_{(i,j)\in S_k} \hat{U}_i\hat{U}_j$$

normalized by a constant. In the time series case which considers regularly spaced data points, the normalizing constants for the covariance at lag k are either the total number N of sample points or the exact number N_k, $N_k=N-|k|$, of terms in the sum estimating the covariance. The first estimate is negatively biased but guarantees that the estimated covariance function is non-negative definite while the second estimate is unbiased (in the case of known mean) but does not ensure a non-negative definite matrix estimate. In the spatial context and particularly for irregularly spaced points, the bias associated with an equivalent form of the first definition would be much amplified and hence the usual estimate for $C(k)$

$$\hat{C}(k)=N_k^{-1}\sum_{(i,j)\in S_k} \hat{U}_i\hat{U}_j,$$

is preferred. Some $\hat{C}(k)$ corresponding to large distances might have to be set to zero to ensure that the matrix $\hat{\Sigma}_D$, with elements $\{\hat{\sigma}_{ij}, (i,j)\in A\times A\}$ defined by

$$\hat{\sigma}_{ij}=\hat{C}(k) \quad \text{if} \quad (i,j)\in S_k \tag{8}$$

is non-negative definite.

Direct estimation of Σ was advocated by Arora & Brown (1977) in a case where spatial data at different time intervals were available, thus allowing an estimation of all the σ_{ij} without the need for further assumptions. In the context of trend surface analysis, Haining (1987) has also used direct estimation of Σ defining the strata S_k as increasing order of neighbourhoods with respect to a contiguity matrix.

2.2 *Estimation of the regression coefficients*

2.2.1 *Maximum likelihood estimation*. In the case where Σ in (1) is modelled in terms of a vector of parameters θ, the profile likelihood (2), which is only dependent on θ, can be maximized numerically, giving the maximum likelihood estimate $\hat{\theta}$ of θ and an estimation of $\hat{\Sigma}=\Sigma(\hat{\theta})$. Conditionally on $\hat{\Sigma}$, we then obtain:

$$\hat{\beta}=(X^t\hat{\Sigma}^{-1}X)^{-1}X^t\hat{\Sigma}^{-1}Y. \tag{9}$$

Writing $\Sigma=\sigma^2 D$, we further evaluate $\hat{\sigma}^2=N^{-1}(Y-X\hat{\beta})^t\hat{D}^{-1}(Y-X\hat{\beta})$ and the matrix $(X^t\hat{D}^{-1}X)^{-1}\hat{\sigma}^2$ as the estimated asymptotic variance–covariance matrix of $\hat{\beta}$. From these formulae approximate tests of the regression coefficients can be performed. Mardia & Marshall (1984) have studied the asymptotic properties of the maximum likelihood estimators $(\hat{\beta},\hat{\theta})$. Assuming that Y is a Gaussian process, they give conditions that ensure the consistency and the asymptotic normality of $(\hat{\beta},\hat{\theta})$. In a simulation study aimed at assessing qualitatively the small sample properties of the maximum likelihood estimators, Mardia & Marshall noticed that when θ is composed of a variance parameter σ^2 and a shape parameter α for the covariance, then $\hat{\sigma}^2$ is negatively biased. This might imply an underestimation of the standard errors of $\hat{\beta}$, the more so since these standard errors are also estimated conditionally on the estimated $\hat{\alpha}$. This had led some authors to consider a Bayesian approach (Hepple, 1979).

When the data set contains a large number of points, maximum likelihood estimation becomes computationally very heavy and is fraught with difficulties. Ripley (1988) shows that the profile likelihood of the SAR and CAR models with constant mean is unimodal but not necessarily concave. For other covariance schemes the profile likelihood might be multimodal (Warnes & Ripley, 1987). A further study by Mardia & Watkins (1989) indicates that for small samples, multimodality can be quite frequent for a number of covariance schemes and that for the spherical scheme (three-dimensional analogue of the disc model) this phenomenon is also increased with N possibly in connection with the lack of a second derivative with respect to the range parameter for this scheme. They recommend a graphical study of the profile likelihood $L_p(\theta, Y)$. $L_p(\theta, Y)$ can easily be drawn in the cases that they have considered, i.e. where the shape of the covariance is modelled in terms of a single parameter, but a study of $L_p(\theta, Y)$ becomes computationally and graphically more involved when there is more than one shape parameter.

When the number of sample points is too large, the maximum likelihood procedure becomes computationally prohibitive since Σ^{-1} has to be evaluated. For stationary processes defined on a lattice, it is possible to approximate the likelihood (see Whittle, 1954; Guyon, 1982 and the discussion in Mardia & Marshall, 1984). For non-lattice data, the problem was recently considered by Vecchia (1988) who proposed carrying out estimation within the class of covariance functions with rational spectral density with the help of successive approximate likelihood functions which are easier to handle. He applied his method successfully to simulated data sets and to water level data where he wanted to estimate a trend.

2.2.2 Iterative EGLS. Another possible estimation method takes the form of an iterative least squares scheme, similar to that proposed by Cochrane & Orcutt (1949) for time series. For spatial regressions, it was first discussed by Ord (1975). We shall consider it only in the framework of direct estimation of Σ even though it could also be used with parametric modelling as is the case in time series (Judge *et al.*, 1985).

At a first step, β is estimated by OLS giving $\hat{\beta}_{\mathrm{OLS}}$. The residuals $\hat{U}_0 = Y - X\hat{\beta}_{\mathrm{OLS}}$ are then computed as well as their autocovariances $\{\hat{C}(k)\}$ and used to give a first estimate of Σ_{D} as defined in (8) corresponding to a fixed choice of strata S_k, $0 \leqslant k \leqslant K$.

In most cases this first estimate is not non-negative definite and some $\hat{C}(k)$ are progressively set equal to zero by decreasing order of distances until a first estimate of Σ_{D}, $\hat{\Sigma}_{\mathrm{D}}^1$ is reached which is non negative definite.

At a second step, β is re-estimated by generalized least squares, giving the two-steps estimator $\hat{\beta}_1 : \hat{\beta}_1 = [X^t(\hat{\Sigma}_{\mathrm{D}}^1)^{-1}X]^{-1}X^t(\hat{\Sigma}_{\mathrm{D}}^1)^{-1}Y$ and new residual $\hat{U}_1 = Y - X\hat{\beta}_1$ are calculated. The whole cycle is then repeated with $\hat{U}_1$ instead of $\hat{U}_0$ in the first step, giving successively $\hat{\beta}_i$ and $\hat{\Sigma}_{\mathrm{D}}^i$, $i \geqslant 1$, until convergence. Convergence is considered to have been reached when the standardized estimates of $\hat{\beta}_i$ differ only in the fourth decimal place. When an estimated matrix $\hat{\Sigma}_{\mathrm{D}}^i$ which is not non-negative definite is obtained during the iterations, the whole cycle is started anew with a decreased initial number of non-zero autocorrelations. The final number of non-zero $\hat{C}(k)$ used in $\hat{\Sigma}_{\mathrm{D}}^i$ is recorded as L.

2.3 Relative efficiency of OLS with regard to GLS

General results on the efficiency of OLS and GLS procedures are not available since it will depend in a complex way both on X and Σ. For time series regression, Grenander (1954) expressed the asymptotic efficiency of the OLS estimate in terms of the spectral distribution of the regression sequence and the spectral density of the error process. As a corollary, he pointed out that in the case of polynomial or trigonometric regression, the OLS are asymptotically efficient but that this will not hold in general for regressors obtained as measurements of certain variables which are of a non-deterministic nature as in our examples. The relative efficiency of OLS and GLS estimates has also been investigated in the time series case. For example, Fuller (1976) derives the limiting ratio of the variances of the OLS and GLS estimates, when both X (one-dimensional) and U are AR(1) processes, as a rational function of both autocorrelation parameters.

In a recent paper, Krämer & Donninger (1987) have given some results on the relative efficiency of OLS with regard to GLS for the simultaneous spatial autoregressive model (3). They define a relative efficiency, e, by the quotient of the traces of the covariance matrices for the GLS (which is also the maximum likelihood estimate when Σ is fixed) and OLS coefficient estimates, respectively:

$$e=\operatorname{tr}(X^t\Sigma^{-1}X)^{-1}/\operatorname{tr}[(X^tX)^{-1}X^t\Sigma X(X^tX)^{-1}] \tag{10}$$

They show that for $\Sigma=\sigma^2(I-bW)^{-1}(I-bW^t)^{-1}$, the limit of e is unity as b tends to its maximal value of 1 provided the regressors $\{X\}$ include a constant term. As e is also equal to 1 when $b=0$, this shows that the relative efficiency has a bath tub shape in this particular case. Indeed Krämer & Donninger point out that for intermediate values of b, which is the case of most interest, the loss of efficiency can be substantial. As Dielman & Pfaffenberger (1989) have commented, if the variance of the estimated intercept is large with respect to the variances of the slope coefficients, this global measure of relative efficiency might mask the poor performance of the OLS with regard to GLS specifically on the slope coefficients, which are in practise the parameters of interest. In order to pursue this problem further, we have computed both the global measure e and the ratio of the variances of $\hat{\beta}$ (defined in (9)) and of $\hat{\beta}_{OLS}$ for some slope coefficients of interest in our examples.

2.4 Measures of fit

There is no single recognized measure of fit to help discriminate between the different spatial parametric forms (3)–(7) since they define non-nested models. Assuming isotropy, graphical methods such as the drawing of the variogram or correlogram of the OLS residuals $\hat{U}_0=Y-X\hat{\beta}_{OLS}$, and the fitting or the checking of a functional form for these have been suggested (Ripley, 1981; Cook & Pocock, 1983). This could be misleading as the variance–covariance matrix of $\hat{U}_0$ is different from Σ:

$$\operatorname{Var}(\hat{U}_0)=\Sigma-\Sigma H-H\Sigma+H\Sigma H,$$

where H is the hat matrix, $H=X(X^tX)^{-1}X^t$. In our examples, we have nevertheless found a parallel form of decrease in Σ and in $\operatorname{Var}(\hat{U}_0)$ of the correlations as a function of distance. Note also that the shape of the estimated variogram or correlogram can be strongly influenced by the choice of distance classes.

Global improvement provided by the spatial modelling in comparison to the non spatial regression with $\Sigma=\sigma^2 I$ can be tested by a χ^2 test of improvement of the likelihood. However, parameter estimates are not, in most cases, directly comparable between models even if the models share the same dimension.

It has been argued by some authors that model choice could be based on predictive ability (Box, 1980; Gelfand *et al.*, 1991). In particular the prediction sum of squares or other checking functions provide a measure of fit that can be directly compared between models. For model (1) we define the prediction residual for observation i as $e_{p,i}$:

$$e_{p,i}=(\sigma^{ii})^{-1}\left(\sum_{j\in A}\sigma^{ij}e_j\right),$$

where (σ^{ij}) is the (i,j) element of Σ^{-1}, e_j is the jth deviation, $e_j = Y(j) - x_j^t\hat{\beta}$, with x_j representing the jth column of X^t and $\hat{\beta}$ is estimated as in (9). The prediction sum of squares (PRESS) is then equal to

$$\left(\sum_{i\in A} e_{p,i}^2\right).$$

2.5 *Modified tests of correlation and partial correlations for spatially autocorrelated variables*

A modified test of correlation between spatial variables has been proposed by Clifford *et al.* (1989) and extended to tests of partial correlations by Richardson (1990). It is based on an estimation of the variance of the correlation or partial correlation coefficient which takes into account the internal autocorrelation of the variables.

Supposing first that we are only concerned with testing the correlation r_{XY} between two variables X and Y observed on the domain A. An estimate of the variance of r_{XY} is then calculated as:

$$\widehat{\mathrm{Var}(r_{XY})} = \sum_k N_k \hat{C}_X(k)\hat{C}_Y(k)/N^2 s_X^2 s_Y^2$$

where $\hat{C}_X(k)$ and $\hat{C}_Y(k)$ are, respectively, the autocovariances of X and Y in the strata S_k defined earlier, N_k is the number of pairs in S_k and s_X^2 or s_Y^2 are the respective empirical variances of X and Y over A (Clifford *et al.*, 1989).

An estimated effective sample size $\hat{M}$ is then defined by the relationship:

$$\hat{M} = 1 + (\widehat{\mathrm{Var}\, r_{XY}})^{-1}$$

and a modified t-test of r_{XY} is defined which rejects the null hypothesis of no association when

$$|(\hat{M}-2)^{1/2} r_{XY}(1-r_{XY}^2)^{-1/2}| > t_{\hat{M}-2}^{\alpha} \tag{11}$$

where $t_{\hat{M}-2}^{\alpha}$ is the critical value of the t-statistic with $\hat{M}-2$ degrees of freedom (d.f.)

In the case of several variables, the partial correlation between X and Y conditionally on $Z_1, \ldots, Z_p$ is tested in a similar way (Richardson, 1990). First X and Y are regressed separately on $Z_1, \ldots, Z_p$ by OLS giving estimated residuals $\hat{V}_1$ and $\hat{V}_2$ and secondly the correlation between $\hat{V}_1$ and $\hat{V}_2$ is tested using the modified $t_{\hat{M}-2}$ statistic given in (11) with:

$$\hat{M} = 1 + [\widehat{\mathrm{Var}\, r_{\hat{V}_1\hat{V}_2}}]^{-1}$$

3 Results

3.1 *The data*

Male lung cancer mortality rate has been standardized over the age range 35–74 and over a 2 year period, 1968–1969. The data were provided by the French National Institute for Health and Medical Research (INSERM) at the scale of the French *départements*. Demographic data on the percentage of employed males in specific types of industry were taken from the 1962 census (INSEE). Cigarette sales data were compiled by the French Nationalized Tobacco Company (SEITA). To take into account the time lag between smoking and the onset of a lung pathology, cigarette sales per inhabitant were recorded in 1953. After the grouping of the *départements* around Paris into one area and the exclusion of four others owing to the poor quality of the data, $N = 82$ locations were retained. Among all the types of industry, metal industry, general engineering and textile were found to be significantly linked with lung cancer in a multiple regression analysis assuming no spatial structure for the errors. The same is also true when cigarette sales are included in the multiple regression. Hence we chose regression models including these three industries for our comparative analysis.

3.2 Spatial autocorrelation

All the variables considered exhibit a spatial structure. To quantify this simply, we can calculate Moran autocorrelation coefficient I (see, e.g. Cliff & Ord, 1981). The neighbourhood matrix $W=(W_{ij})$ of weights used in Moran's I was constructed by defining $W_{ij}=1$ if the areas i and j shared a common border and then standardizing W so that its rows sums equal one. Lung cancer, cigarette sales and metal industry are highly autocorrelated with statistically significant autocorrelation coefficients, respectively equal to 0·53, 0·58 and 0·41 (see Richardson, 1990, for a plot of the variograms of some of these variables). The autocorrelation coefficient of textile industry (0·26) is smaller but still significant at the 5% level. Only for general engineering is there no evidence of a spatial structure ($I=0{\cdot}12$, NS). In contrast to the other variables, textile industry has a patchy structure with most of the activity concentrated in the north-east of France and the Rhone Valley area and more than half of the departments showing very small levels of activity. For the other variables, the spatial distribution is much smoother.

Autocorrelation in the residuals of the OLS regression was tested using the generalized Moran coefficient (GMC) (Ripley, 1981) which is constructed in a similar way to I except that the residuals mean is not subtracted. The standard formulae used to derive the moments of the Moran autocorrelation coefficients I are no longer applicable and have been adapted to take into account that the residuals have been estimated. Significant autocorrelation was found in the residuals when the regression model includes the three industries (GMC$=3{\cdot}79$, $p=2\times10^{-4}$). When cigarette sales are further included the residual autocorrelation is weakened but still significant (GMC$=2{\cdot}29$, $p=2\times10^{-2}$).

3.3 Fit of the spatial models

Table 1 presents χ^2 tests of improvement of the likelihood between the standard regression ($\Sigma=\sigma^2 I$) and the regressions performed with the five different parametric models (3)–(7) for Σ as well as the prediction sum of squares. When cigarette sales are not included in the regression (regression A), all the spatial models provide a clearly significant improvement of the likelihood with respect to the standard non-spatial case, thus justifying the need for some spatial modelling of the residuals. There is a notable variation in the size and significance of the improvement among models with the disc model showing a much smaller improvement than the other ones. In contrast, the PRESS

Table 1. Improvement of likelihood[a] and prediction sum of squares under different spatial models[b] for the residuals variance–covariance matrix in the multiple regression of male lung cancer mortality rates on the three industries without (regression A) or with (regression B) inclusion of cigarette sales

	Regression A			Regression B		
Model	χ^2	p	PRESS[c]	χ^2	p	PRESS[c]
OLS	—	—	$1{\cdot}90\times10^{-6}$	—	—	$0{\cdot}96\times10^{-6}$
CAR	$\chi_1^2=25{\cdot}6$	$p<10^{-5}$	$1{\cdot}44\times10^{-6}$	$\chi_1^2=2{\cdot}1$	$p=0{\cdot}15$	$0{\cdot}94\times10^{-6}$
SAR	$\chi_1^2=32{\cdot}0$	$p<10^{-7}$	$1{\cdot}37\times10^{-6}$	$\chi_1^2=8{\cdot}2$	$p=4\times10^{-3}$	$0{\cdot}87\times10^{-6}$
DISC	$\chi_1^2=9{\cdot}7$	$p=2\times10^{-3}$	$1{\cdot}40\times10^{-6}$	$\chi_1^2=9{\cdot}4$	$p=2\times10^{-3}$	$0{\cdot}88\times10^{-6}$
EXPO	$\chi_2^2=30{\cdot}4$	$p<10^{-6}$	$1{\cdot}34\times10^{-6}$	$\chi_2^2=8{\cdot}4$	$p=1{\cdot}5\times10^{-2}$	$0{\cdot}86\times10^{-6}$
BESSEL	$\chi_2^2=25{\cdot}5$	$p=5\times10^{-5}$	$1{\cdot}41\times10^{-6}$	$\chi_2^2=8{\cdot}6$	$p=1{\cdot}4\times10^{-2}$	$0{\cdot}86\times10^{-6}$

[a] Improvement of the likelihood between the standard regression $\Sigma=\sigma^2 I$ and the five spatial parametrizations of Σ.

[b] The five spatial models are defined in equations (3)–(7).

[c] The PRESS criterion is defined in Section 2.4.

criterion, which is reduced by about 30% between the OLS case and the spatial models, exhibits only a small degree of variation between the models. When cigarette sales are included in the regression (regression B), there is less residual autocorrelation and the χ^2 statistics are on the whole substantially decreased in comparison to regression A. The CAR model does not fit very well with a non-significant χ^2 test and a PRESS value nearly equal to that of the OLS. The values of the PRESS criterion for the other models are extremely close.

Table 2 summarizes the maximum likelihood estimates of the spatial parameters involved in the five spatial models (3)–(7) for Σ. For calculating their standard errors we followed the method indicated by Mardia (1990), and also used explicit expressions for the CAR and SAR models in terms of the eigenvalues of W (see details in the Appendix). For the sake of computational simplicity, the standard error of the estimated spatial parameter $\hat{\delta}$ in the Bessel model was calculated conditionally on the estimated value $\hat{\nu}$ of the shape parameter for which we did not derive a standard error.

One has to be careful in interpreting the coefficients in Table 2 as they are not comparable between models. All spatial parameters are clearly significant in regression A and, as expected, indicate weaker spatial dependence in regression B. Note that in regression A the parameter $\hat{c}$ of the CAR model is very close to its maximal permissible value of 0·179 which corresponds to the inverse of the maximal eigenvalue of W. In contrast $\hat{c}$ is low and not significant for regression B. Due to the normalization condition of row sums equal to 1 imposed on the W matrix of the SAR model, the parameter $\hat{b}$ of this model can be interpreted as an overall weighting of the influence of the neighbours. The parameter $\hat{a}$ of the disc model is also directly interpretable, $2\hat{a}$ being the distance threshold (in kilometres) after which the model has estimated that there is no spatial dependence left. Finally the shape parameter of the Bessel model indicates a steeper decrease than the exponential model (which would correspond to $\nu = 1/2$) and a better precision in the estimation of the rate spatial decrease $\hat{\delta}$ than for the corresponding parameter $\hat{\lambda}$ of the exponential model. In summary one could say that this table shows various ways characterizing the spatial dependence but gives only limited information on which to base a choice of model.

3.4 *Comparative regression results for different parametric forms of* Σ

The top six lines of Table 3 (*a*–*c*) give maximum likelihood estimates of regression coefficients for the three types of industry as well as an estimation of the standard error

Table 2. Maximum likelihood estimates of the spatial parameters for the five parametric spatial models[a] for the residuals variance–covariance matrix in the multiple regression of male lung cancer mortality rates on the three industries without (regression A) or with (regression B) inclusion of cigarette sales

	Regression A		Regression B	
Model	parameter(s)	(se)[b]	parameter(s)	(se)[b]
CAR	$\hat{c}=$ 0·175	(0·005)	$\hat{c}=$ 0·077	(0·057)
SAR	$\hat{b}=$ 0·613	(0·103)	$\hat{b}=$ 0·353	(0·138)
DISC	$\hat{a}=$ 94·31	(5·66)	$\hat{a}=$ 35·73	(4·92)
EXP	$\hat{\gamma}=$ 0·745	(0·091)	$\hat{\gamma}=$ 0·554	(0·265)
	$\hat{\lambda}=$ 0·0035	(0·0018)	$\hat{\lambda}=$ 0·0119	(0·0083)
BESSEL	$\hat{\nu}=$ 0·305		$\hat{\nu}=$ 0·228	
	$\hat{\delta}=$ 0·0089	(0·0013)	$\hat{\delta}=$ 0·0133	(0·0042)

[a] The five spatial models are defined in equations (3)–(7).
[b] For the calculation of the standard error, see the Appendix.

of $\hat{\beta}$ (see equation (9) and the formulae which follow). Overall the different spatial modelling of the residuals have little influence on the standard errors of the regression slopes whereas there is a manifest influence on the slope estimates themselves. With respect to the slope estimates, the effect is quite different for the three industries.

For the metal industry, the spatial modelling of the residuals has nearly halved the slope estimates and the associated t-statistics in regression A. All models find a significant association but there is quite a range of significance levels. Results are closer between the CAR and disc models, on one hand, and the SAR, exponential and Bessel models, on the other hand. The same pattern is also apparent in regression B even though the differences between the standard and the spatial regression results have weakened.

In contrast, for general engineering, the spatial parameterization has hardly changed the slope estimates whether in regression A or B. There is overall similarity of the slopes for all models, the only difference is a small increase of the slope for the SAR model. The associated t-statistics are on the whole slightly larger than without spatial modelling. For the textile industry, the spatial parameterizations have substantially reduced the slope estimates in regression A and all the models find a non-significant association. In regression B, all the statistics have increased but there is evidence of a discrepancy between

Table 3. Results from the multiple regression of male lung cancer mortality rates on the three industries given by different methods of analysis

Model	Regression A $\hat{\beta}$ (10^{-3})	$\hat{\sigma}_{\hat{\beta}}$ (10^{-3})	t	p	Regression B $\hat{\beta}$ (10^{-3})	$\hat{\sigma}_{\hat{\beta}}$ (10^{-3})	t	p
(a) Results for the metal industry								
OLS	2·46	0·39	6·15	$\leqslant 10^{-7}$	1·50	0·27	4·87	$\leqslant 10^{-5}$
CAR	1·75	0·38	4·47	$\leqslant 10^{-4}$	1·35	0·30	4·31	$\leqslant 10^{-4}$
SAR	1·39	0·42	3·25	$1{\cdot}8 \times 10^{-3}$	1·11	0·32	3·37	$1{\cdot}2 \times 10^{-3}$
DISC	1·63	0·47	3·80	$0{\cdot}3 \times 10^{-3}$	1·25	0·30	3·99	$0{\cdot}2 \times 10^{-3}$
EXPO	1·28	0·43	2·87	$5{\cdot}3 \times 10^{-3}$	1·01	0·32	3·02	$3{\cdot}5 \times 10^{-3}$
BESSEL	1·35	0·43	3·02	$3{\cdot}5 \times 10^{-3}$	1·05	0·32	3·17	$2{\cdot}2 \times 10^{-3}$
EGLS 2 Steps	2·15	0·37	5·72	$\leqslant 10^{-6}$	1·32	0·29	4·38	$\leqslant 10^{-4}$
EGLS Iterative	2·03	0·38	5·27	$\leqslant 10^{-5}$	1·23	0·31	4·04	$0{\cdot}1 \times 10^{-3}$
$t_{\hat{M}-2}$	$\hat{M}=31$		3·77	$0{\cdot}8 \times 10^{-3}$	$\hat{M}=50$		3·86	$0{\cdot}4 \times 10^{-3}$
(b) Results for general engineering								
OLS	1·88	0·63	2·87	$5{\cdot}3 \times 10^{-3}$	1·29	0·46	2·74	$7{\cdot}8 \times 10^{-3}$
CAR	1·90	0·57	3·25	$1{\cdot}8 \times 10^{-3}$	1·25	0·46	2·63	$10{\cdot}4 \times 10^{-3}$
SAR	2·11	0·55	3·74	$0{\cdot}4 \times 10^{-3}$	1·37	0·45	2·96	$4{\cdot}1 \times 10^{-3}$
DISC	1·86	0·52	3·50	$0{\cdot}8 \times 10^{-3}$	1·37	0·44	3·03	$3{\cdot}4 \times 10^{-3}$
EXPO	1·87	0·55	3·32	$1{\cdot}4 \times 10^{-3}$	1·23	0·44	2·69	$8{\cdot}8 \times 10^{-3}$
BESSEL	1·90	0·57	3·28	$1{\cdot}6 \times 10^{-3}$	1·24	0·44	2·70	$8{\cdot}6 \times 10^{-3}$
EGLS 2 Steps	1·72	0·54	3·17	$2{\cdot}2 \times 10^{-3}$	1·01	0·41	2·41	$18{\cdot}6 \times 10^{-3}$
EGLS Iterative	1·89	0·54	3·47	$0{\cdot}9 \times 10^{-3}$	0·89	0·42	2·22	$29{\cdot}7 \times 10^{-3}$
$t_{\hat{M}-2}$	$\hat{M}=71$		2·72	$8{\cdot}3 \times 10^{-3}$	$\hat{M}=69$		2·57	$12{\cdot}4 \times 10^{-3}$
(c) Results for the textile industry								
OLS	1·12	0·53	2·03	$4{\cdot}6 \times 10^{-2}$	0·84	0·38	2·11	$3{\cdot}8 \times 10^{-2}$
CAR	0·52	0·50	1·02	0·32	0·77	0·39	1·93	$5{\cdot}7 \times 10^{-2}$
SAR	0·38	0·49	0·76	0·45	0·61	0·39	1·53	0·14
DISC	0·24	0·47	0·49	0·63	0·84	0·37	2·19	$3{\cdot}2 \times 10^{-2}$
EXPO	0·21	0·49	0·41	0·69	0·58	0·38	1·47	0·15
BESSEL	0·38	0·50	0·74	0·47	0·62	0·38	1·58	0·12
EGLS 2 Steps	0·61	0·46	1·29	0·21	0·88	0·35	2·48	$1{\cdot}5 \times 10^{-2}$
EGLS Iterative	0·55	0·46	1·18	0·25	0·83	0·37	2·47	$1{\cdot}6 \times 10^{-2}$
$t_{\hat{M}-2}$	$\hat{M}=68$		1·88	$6{\cdot}5 \times 10^{-2}$	$\hat{M}=79$		2·12	$3{\cdot}7 \times 10^{-2}$

the models with t-statistics which are either non-significant, borderline or significant at the 3% level (disc model).

Thus we have found that the choice of parametric model can influence quite substantially the slope estimates for two of the three industries.

Figure 1 illustrates some of the difficulties in ascertaining the parametric form of the model by inspection of the graphs of the correlograms as has often been suggested. It represents the correlogram of the OLS residuals together with those corresponding to the five parametric models (3)–(7) for regression A. Parameters have been replaced by their maximum likelihood estimates for the estimation of the correlograms of the five models. For the disc, exponential and Bessel models, the form of the correlogram as a function of distance has been directly represented. For the CAR and SAR models, the autocorrelation for increasing neighbourhood orders defined by a discrete distance function has been plotted against the average distance of all pairs of sites in that neighbourhood class.

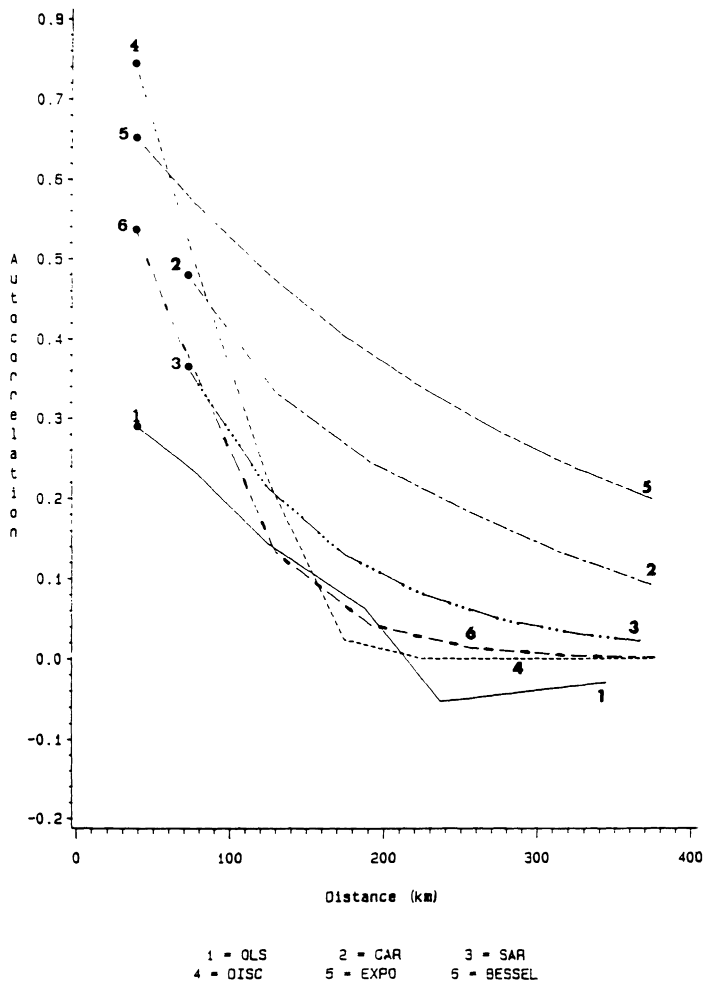

Fig. 1. Correlograms of the residuals from the multiple regression of male lung cancer mortality rates on the three industries (regression A) estimated either by OLS or with a spatially parametrized variance–covariance matrix.

The correlogram of the OLS residuals exhibits a slow, fairly linear decrease and has a downward bias as expected. One can see on Fig. 1 that the disc and the exponential models stand out. The disc model assumes a very steep decrease of the autocorrelation for which there is no evidence on the estimated correlogram of the OLS residuals. In contrast the exponential model assumes a very slow decrease as if it was influenced by some trend in the data. The Bessel model exhibits a steep decrease for small distances but a shape similar to that of the SAR model for larger distances. One can see that no model manages to reproduce the fairly linear decrease observed at the start of the OLS correlogram and that no strong indication emerges from this comparison which could help to choose a particular parametric form except may be to exclude the disc model.

The model correlograms corresponding to regression B (figure not shown) are more steeply decreasing for small distances than those shown on Fig. 1 but the respective positions for the five models are similar to Fig. 1.

3.5 Likelihood profile

We encountered evidence of multimodality in the likelihoods only in the case of the disc model. This can be related to the multimodality behaviour of the spherical model found by

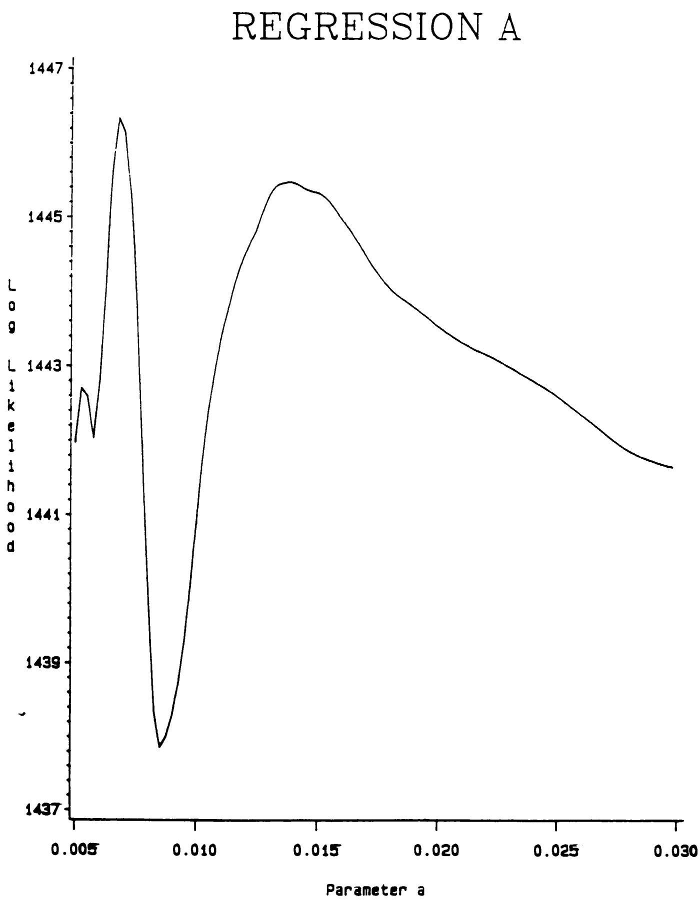

Fig. 2. Likelihood profile (regression A) when the residuals variance–covariance matrix is parametrized by a disc model.

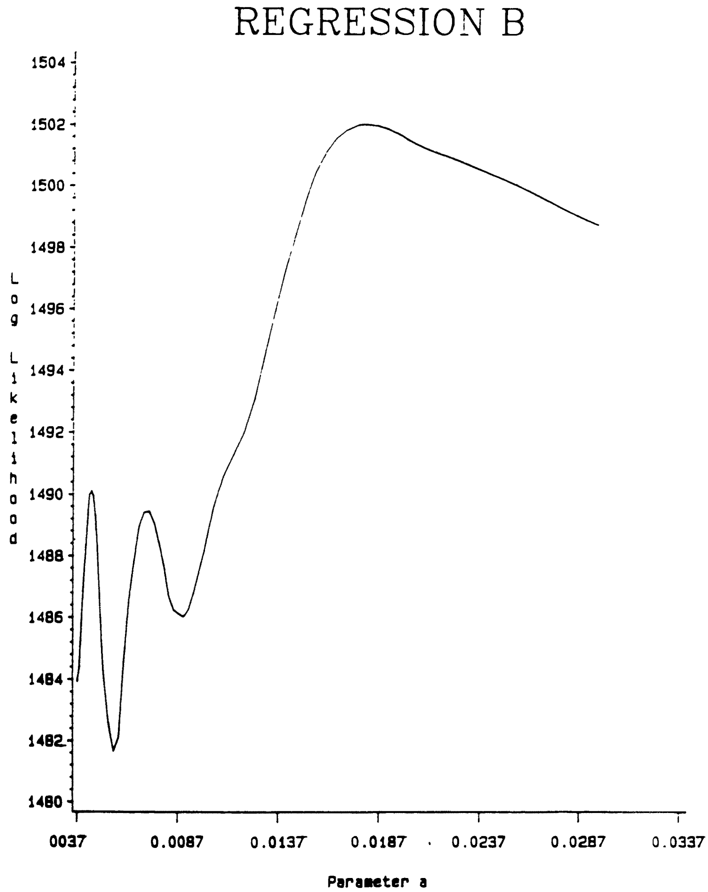

Fig. 3. Likelihood profile (regression B) when the residuals variance–covariance matrix is parametrized by a disc model.

Mardia & Watkins (1989) in their simulation study. The profile likelihood of the disc model is clearly multimodal for both regressions A and B. For regression A (Fig. 2) it is the second peak (corresponding to a high autocorrelation) which maximizes the likelihood whereas in regression B (Fig. 3) it is the third peak. This points to the first peaks influenced by the long-range structure of the data. As shown in Ripley (1988), the profile likelihood of the CAR model is strongly asymmetric and the maximum was obtained for a parameter value equal to 0·175, which is close to the maximal permissible value of 0·179. Hence, the search for a maximum requires some care in this case. In contrast, the profile likelihood of the SAR model is fairly symmetrical with no boundary problems.

3.6. Comparison with modified $t_{\hat{M}-2}$ *test*

The different spatial structures for the three industries are well reflected in the estimated effective sample size $\hat{M}$ of the modified $t_{\hat{M}-2}$ tests (Table 3). The sample size is much reduced for the highly autocorrelated metal industry (regression A) and subsequently increased when cigarette sales are introduced as the residual autocorrelation is weakened. For general engineering and the textile industry, the effective sample size is close to the original one ($N=82$).

On the whole, the significance levels given by the modified tests are within the range of those given by the models for the metal industry and general engineering. For the textile industry, in contrast to all the models, the modified test points to a borderline association for regression A, having found only weak evidence of a spatial structure in the residuals. In regression B where there was a discrepancy between the results given by the different models, the modified test find no evidence of a residual spatial structure with modified d.f. nearly equal to the original sample size and consequently find a significant association with the textile industry similar to the disc and CAR models.

3.7 *Direct estimation of* Σ

As was shown previously, there are some cases where the choice of parametric model for Σ can influence the significance of the regression parameters. Hence, direct estimation of Σ, a procedure that avoids parametric modelling, might be appealing. In implementing this procedure, we are confronted by several problems. The simplest method along these lines is a two-step EGLS method which estimates β by $\hat{\beta}_1$. Its performance will depend both on the number L of the autocovariances used in defining $\hat{\Sigma}_D$ and their estimation. An iterative procedure which tries to minimize the generalized sum of squares in both β and the directly estimated Σ_D seems *a priori* preferable but involves numerical difficulties. In many instances when L is large, we found that one of the directly estimated matrices during the iterations was not non negative definite and that the whole cycle had to be started again with a decreased L. Hence in the regression A, L diminishes from 11 to 6 to obtain convergence in the full iterative process whereas the two-step EGLS can be estimated with $L=11$. For both regressions, convergence was obtained in less than 10 iterative cycles.

Overall estimation and tests of regression slopes are similar for the two-step and for the iterative EGLS method (Table 3). For the metal industry, the results from the direct estimation procedures stand apart in comparison with the parametric models. For this highly autocorrelated variable. EGLS procedures do not adjust enough and tests of regression coefficients seem over significant. For general engineering, which is only weakly autocorrelated, EGLS performs similarly to parametric modelling. For the textile industry (regression B) where there was a discrepancy between the regression results under different parametric models, EGLS finds a stronger association than the OLS regression.

The apparent inability of EGLS to give correct significance levels with high autocorrelation was checked in a simulation study. Pairs of mutually independent normally distributed random variables (X, Y) were simulated on an irregular domain A. The spatial structure of X and Y on A was generated by the disc model. The domain A was chosen to be similar to that of the analysed examples (see Richardson, 1990, for details on an equivalent simulation model). Parameters of the disc model were chosen so that the autocorrelation for a distance of 40 km between points, $\rho(1)$ would be equal to 0·6 or 0·8 for both X and Y. Tests of the regression coefficients were performed at a nominal level of 5% but observed rejection levels for the iterative EGLS method were much higher. The observed proportion of type I errors for 1000 simulations carried out with $\rho(1)=0{\cdot}6$ was 18·8%, 95% CI [16·4%; 21·1%] and that for 1000 simulations carried out with $\rho(1)=0{\cdot}8$ was 38·1%, 95% CI [35·1%; 41·1%]. Both results show that the EGLS method will lead to over significant tests in cases of strong positive autocorrelation in the variables X and Y.

3.8 *Relative efficiency*

Table 4 displays the OLS variance of the slope coefficients $\hat{\beta}_{OLS}$ for two of the industries studied as well as an estimation of the variances of $\hat{\beta}_{OLS}$ and $\hat{\beta}$ as given in (9) when either spatial models (3)–(7) are assumed for the residuals or iterative EGLS was carried out.

Table 4. Relative efficiency of regression parameters between spatial and non spatial regression models

Model	Var ($\hat{\beta}$)[a] (10^{-7})	Var ($\hat{\beta}_{OLS}$)[b] (10^{-7})	Ratio[c]	e[d]
(*a*) *Metal industry*				
OLS	1·5	1·5	1	1
CAR	1·5	2·2	0·66	0·52
SAR	1·7	3·1	0·56	0·58
DISC	2·2	6·4	0·34	0·30
EXPO	1·9	5·3	0·35	0·40
BESSEL	1·9	3·4	0·55	0·59
EGLS iterative	1·4	2·3	0·62	0·62
(*b*) *General engineering*				
OLS	4·0	4·0	1	1
CAR	3·3	7·7	0·43	0·52
SAR	3·0	5·2	0·58	0·58
DISC	2·7	9·9	0·27	0·30
EXPO	3·0	5·7	0·53	0·40
BESSEL	3·2	5·4	0·60	0·59
EGLS iterative	2·9	4·9	0·59	0·62

[a] $\text{Var}(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}$.
[b] $\text{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$.
[c] Ratio = Var ($\hat{\beta}$)/Var ($\hat{\beta}_{OLS}$).
[d] Defined in equation (10).

These variances are, respectively, equal to:

$$\text{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1},$$

$$\text{Var}(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}.$$

To estimate them, Σ is replaced by $\Sigma(\hat{\theta}_{M.L.})$ for models (3)–(7) and by $\hat{\Sigma}_D$ for the iterative EGLS. The third column of Table 4 calculates the ratio of these variances specifically for the metal industry or general engineering while the last column displays the trace efficiency criterion e defined in (10). Two features stand out.

The OLS variance estimate, $(X'X)^{-1}$, clearly underestimates the true variance of $\hat{\beta}_{OLS}$ (second column). Furthermore, the variance of $\hat{\beta}_{OLS}$ is strongly dependent on the spatial model assumed for the error structure and nearly trebles for the metal industry between the CAR and the disc model. Since the true model of the error structure will not be known, this highlights one of the disturbing features resulting from using $\hat{\beta}_{OLS}$ instead of the maximum likelihood estimate $\hat{\beta}$. In contrast the variances of $\hat{\beta}$ estimated by maximum likelihood (first column) show much less variation between models and are always substantially smaller than the corresponding OLS variances. The loss of efficiency shown can be quite substantial but is model dependent. However, we found little difference between the global relative efficiency e based on the trace criterion defined by Krämer & Donninger (1987) and the individual relative efficiency for each slope coefficient (third column). Note also that the variances corresponding to the iterative EGLS method are always amongst the smallest given by the models. This is related to the insufficient adjustment for spatial structure given by this method.

Finally, it is interesting to note that the OLS variance estimates, $(X'X)^{-1}$, are quite close to the variance estimates of $\hat{\beta}$ estimated by maximum likelihood (see also Table 3). Consequently using the OLS variance would not have greatly changed the significance of the tests of the slope estimates if these estimates had not themselves varied and in particular decreased between OLS and the maximum likelihood estimation. This phenomenon was also noticed by Cook & Pocock (1983) in their study of cardiovascular mortality and water hardness.

4 Discussion

Overall in this analysis, we have found a geographical association between lung cancer mortality for men, cigarette sales and two industries, metal and general engineering. These results are consistent with those of epidemiological studies where working in the metal industry, ship building or motor vehicle construction has been recognized as presenting a risk for lung cancer due to potential exposure to several carcinogenic substances (Simonato *et al.*, 1988). These two industrial variables, together with cigarette sales, explain a large part of the spatial structure of lung cancer but there is still evidence of a residual spatial structure, which suggest a possible influence of environmental risk factors.

We shall now discuss in turn the influence of the chosen spatial model for the residuals on the estimation and the testing of the regression coefficients, and the choice of model.

Our analysis shows that the agreement or not between the estimated coefficients under different models is closely related to the spatial structure of each variable, with discrepancies arising either in the case of a strongly autocorrelated spatial structure or when the structure is more patchy. Since the modified $t_{\hat{M}-2}$ test adjusts separately for the residual spatial structure of each variable, it is useful to compute it at a first step to have an assessment of the association which is not model dependent and thus potentially signal extreme model effects. There are many problems arising from using an iterative EGLS procedure with non-lattice data, e.g. arbitrary choice of a cut-off lag L, no guarantee of non-negative definiteness of the estimated variance–covariance matrix, inflated type I errors, which seem to indicate that this procedure is not suitable. It would be interesting to further investigate by simulation the robustness of the EGLS in comparison to that of the modified $t_{\hat{M}-2}$ test and parametric modelling for an incorrect model choice.

From a data analytic point of view, it is important to be able to compare the fit of different models in order to have some indication of a 'best choice'. When the models are nested, this can be performed by appropriately testing the log-likelihood differences as done for instance in Cressie & Chan (1989). The problem is more difficult for non-nested models, having possibly different dimensionality.

The fitted correlograms highlight interesting differences between the models but are difficult to interpret for model choice. One can see that the disc model is constrained so that a high autocorrelation for short distances is accompanied by a very sharp decrease. Its multimodality can be intuitively explained by a conflict between parameters compatible with high autocorrelation at short distances and those compatible with the cut-off distance from which autocorrelations are presumed zero. At the other extreme, the exponential model is influenced by the long-term trend in the data. The Bessel model, which requires much more lengthy maximizations than the other models, does not seem to reproduce the shape of the OLS correlogram better than the SAR model.

It has long been known that to estimate the predictive ability of a regression model on the calibration sample on which it was estimated is over optimistic and that shrinkage is usually observed on cross-validation (Stone, 1974). Hence, the prediction sum of squares that we have evaluated is certainly too small in the absolute but nevertheless might still give some useful insight for model comparison. Cross-validation was discussed for the fitting of spatial covariances by Ripley (1981). The basic idea is to delete one observation i at a time and then to predict it using the remaining $N-1$ observations. If the variance–covariance matrix Σ is known, one could use the results given by Martin (1984, 1989) to calculate, for instance, the cross-validation estimate of the predictive residuals, i.e. the difference between the the observed value and that predicted from the remaining $N-1$ observations and form the sum of the squares of those differences or other checking functions. When Σ is not known but is parametrized, $\Sigma = \Sigma(\theta)$, cross-validation would involve carrying out for each i, first, a numerical maximization of the likelihood based on $N-1$ points to determine $\hat{\theta}_{(i)}$ and, secondly, the calculation of the predicted value using $\Sigma(\hat{\theta}_{(i)})$ and the remaining $N-1$ observations. In practice this is unfeasible since it involves

N numerical maximizations is for each model. An approximate criterion could be calculated by performing the first step only once using the N sample points to determine $\hat{\theta}$ and then reasoning as if $\Sigma(\hat{\theta})$ is known. We have not attempted to explore the properties of such a criterion for model choice in spatial regression. It is an interesting area for future research.

Acknowledgments

The authors warmly thank Evelyne Przybilski for secretarial assistance. This was supported by EURATOM contract BI6-126F. The final version of this paper was completed while S.R. was visiting the MRC Biostatistics Unit (Cambridge, UK) supported by an INSERM–MRC exchange fellowship.

References

ABRAMOWITZ, M. & STEGUN, I. (1965) *Handbook of Mathematical Functions* (New York, Dover).

AGTERBERG, F. P. (1984) Trend surface analysis, in E. L. GAILE & C. J. WILMOT (Eds), *Spatial Statistics and Models* (Dordrecht, Reidel).

ANSELIN, L. (1988) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity, *Geographical Analysis*, 88, pp. 1–17.

ARORA, S. S. & BROWN, M. (1977) Alternative approaches to spatial autocorrelation: an improvement over current practice, *International Regional Science Review*, 2, pp. 67–78.

BESAG, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B*, 36, pp. 192–236.

BIVAND, R. S. (1984) Regression modeling with spatial dependence: an application of some class selection and estimation methods, *Geographical Analysis*, 16, pp. 25–37.

BOX, G. (1980). Sampling and Bayes' inference in scientific modelling and robustness (with discussion), *Journal of the Royal Statistical Society, Series A*, 143, pp. 382–430.

CLIFF, A. D. & ORD, J. K. (1981) *Spatial Processes. Models and Applications* (London, Pion).

CLIFFORD, P., RICHARDSON, S. & HÉMON, D. (1989) Assessing the significance of the correlation between two spatial processes, *Biometrics*, 45, pp. 123–134.

COCHRANE, D. & ORCUTT, G. H. (1949) Applications of least-square regressions to relationships containing autocorrelated error terms, *Journal of the American Statistical Association*, 44, pp. 32–61.

COOK, D. G. & POCOCK, S. J. (1983) Multiple regression in geographic mortality studies with allowance for spatially correlated errors, *Biometrics*, 39, pp. 361–371.

CRESSIE, N. & CHAN, N. H. (1989) Spatial modeling of regional variables, *Journal of the American Statistical Association*, 84, pp. 393–401.

DIELMAN, T. E. & PFAFFENBERGER, R. C. (1989) Efficiency of ordinary least squares for linear models with autocorrelation, *Journal of the American Statistical Association*, 84, p. 248.

DOREIAN, P. (1981) 'Estimating linear models with spatially distributed data', in: S. LEINHARDT (Ed.), *Sociological Methodology* (San Francisco, Jossey-Bass).

FULLER, W. A. (1976) *Introduction to Statistical Time Series* (New York, John Wiley).

GELFAND, A. E., DEY, D. K. & CHANG, H. (1991) In: J. BERGER, J. BERNARDO, A. P. DAVID and A. F. M. SMITH, (Eds), *Bayesian Statistics 4: Proceedings of the Fourth International Meeting on Bayesian Statistics* (Oxford, Oxford University Press).

GRENANDER, U. (1954) On the estimation of regression coefficients in the case of an autocorrelated disturbance, *Annals of Mathematical Statistics*, 25, pp. 252–272.

GUYON, X. (1982) Parametric estimation for a stationary process on a d-dimensional lattice, *Biometrika*, 69, pp. 95–105.

HAINING, P. (1987) Trend surface models with regional and local scales of variation with an application to aerial survey data, *Technometrics*, 29, pp. 461–469.

HAINING, R. P. (1990) Models in human geography: problems in specifying, estimating and validating models for spatial data, in: *Institute of Mathematical Geography Monograph Series, Vol. 12, Spatial Statistics: Past, Present, and Future* (Ann Arbor, Institute of Mathematical Geography).

HEPPLE, L. W. (1979) Bayesian analysis of the linear model with spatial dependence, in: C. P. A. BARTELS & R. H. KETELLAPPER (Eds), *Exploratory and Explanatory Statistical Analysis of Spatial Data* (Boston, Martinus Nijhoff).

JUDGE, G., GRIFFITHS, W., HILL, R., LUBKEPOHL, H. & LEE, T. C. (1985) *The Theory and Practice of Econometrics*, 2nd edn (New York, John Wiley).
KRÄMER, W. and DONNINGER, C. (1987) Spatial autocorrelation among errors and the relative efficiency of OLS in the linear regression model, *Journal of the Americal Statistical Association*, 82, pp. 577–579.
MALIN, S. R. C. and HIDE, R. (1982) Bumps on the core–mantle boundary: geomagnetic and gravitational evidence revisited, *Philosophical Transactions of the Royal Society of London, Series A*, 306, pp. 281–289.
MARDIA, K. V. (1990) Maximum likelihood estimation for spatial model, in: *Institute of Mathematical Geography Monograph Series. Vol. 12, Spatial Statistics: Past, Present, and Future* (An Arbor, Institute of Mathematical Geography).
MARDIA, K. V. & MARSHALL, R. J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, 71, pp. 135–146.
MARDIA, K. V. & WATKINS, A. J. (1989) On multimodality of the likelihood in the spatial linear model, *Biometrika*, 76, pp. 289–295.
MARTIN, R. J. (1984) Exact maximum likelihood for incomplete data from a correlated gaussian process, *Communications in Statistics—Theory and Methods*, 13, pp. 1275–1288.
MARTIN, R. J. (1990) The use of time-series models and methods in the analysis of agricultural field trials, *Communications in Statistics—Theory and Methods*, 19, pp. 55–81.
MARTIN, R. J. (1992). Leverage, influence and residuals in regression models when observations are correlated, *Communications in Statistics—Theory and Methods*, 21, pp. 1183–1212.
MOLLIÉ, A. & RICHARDSON, S. (1991) Empirical Bayes estimates of cancer mortality rates using spatial models, *Statistics in Medicine*, 10, pp. 95–112.
ORD, K. (1975) Estimation methods for models of spatial interaction, *Journal of the American Statistical Association*, 70, pp. 120–126.
RICHARDSON, S. (1990) A method for testing the significance of geographical correlations with application to industrial lung cancer in France, *Statistics in Medicine*, 9, pp. 515–528.
RIPLEY, B. D. (1981) *Spatial Statistics* (New York Wiley).
RIPLEY, B. D. (1988) *Statistical Inference for Spatial Processes* (Cambridge, Cambridge University Press).
SIMONATO, L., VINEIS, P. & Fletcher, A. (1988) Estimates of the proportion of lung cancer attributable to occupational exposure, *Carcinogenesis*, 9, pp. 1159–1165.
STONE, M. (1974) Cross-validatory choice and assessment of statistical predictions (with discussion), *Journal of the Royal Statistical Society, Series B*, 36, pp. 111–132.
UPTON, G. & FINGLETON, B. (1985) *Spatial Data Analysis by Example* (Chichester, Wiley).
VECCHIA, A. V. (1988) Estimation and model identification for continuous spatial processes, *Journal of the Royal Statistical Society, Series B*, 50, pp. 297–312.
WARNES, J. J. & RIPLEY, B. D. (1987) Problems with likelihood estimation of covariance functions of spatial Gaussian processes, *Biometrika*, 74, pp. 640–642.
WHITTLE, P. (1954) On stationary processes in the plane, *Biometrika*, 41, pp. 434–449.
ZIMMERMAN, D. L. & ZIMMERMAN, H. B. (1991) A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors, *Technometrics*, 33, pp. 77–91.

Appendix: Derivation of the standard errors of the spatial parameter estimates

We suppose as in equation (1) that

$$Y = X\beta + U$$

where β is a $p \times 1$ vector of parameters, U follows a multinormal distribution $N(0, \Sigma)$ and $\Sigma = \Sigma(\theta)$ is modelled in terms of a q-dimensional vector of parameters $\theta = (\theta_i,\ 1 \leqslant i \leqslant q)$.

Following Mardia & Marshall (1984) and Mardia (1990), the asymptotic covariance matrix of $\hat{\theta}$ is A^{-1}, where a_{ij}, the (i,j) element of the information matrix A, is given by:

$$a_{ij} = \tfrac{1}{2}\mathrm{tr}\,(\Sigma^{-1}\Sigma_i\Sigma^{-1}\Sigma_j)$$

and

$$\Sigma_i = \frac{\partial \Sigma}{\partial \theta_i}.$$

For the disc and the exponential models, the calculation of Σ_i is straightforward. For the Bessel model, we have used the approximation to $\partial/\partial z\ K_\nu(z)$ given in Abramowitz & Stegun (1965) (equation 9.7.4)

$$\frac{\partial}{\partial z} K_\nu(z) \doteqdot \sqrt{\frac{\pi}{2z}}(\exp -z)\left\{1+\frac{\mu+3}{8z}+\frac{(\mu-1)(\mu+15)}{2!(8z)^2}-\frac{(\mu-1)(\mu-9)(\mu+35)}{3!(8z)^3}\right\},$$

where $\mu=4\nu^2$.

In the case of the CAR and SAR processes, the asymptotic variance can be written directly in terms of the eigenvalues $\{\lambda_i\}$ of the weight matrix W.

For the CAR process:

$$\widehat{\mathrm{Var}}(\hat{c})=2\times\left[\sum_{i=1}^{N}\left(\frac{\lambda_i}{1-\hat{c}\lambda_i}\right)^2-\left[\sum_{i=1}^{N}\left(\frac{\lambda_i}{1-\hat{c}\lambda_i}\right)\right]^2\Big/N\right]^{-1}.$$

For the SAR process, the expression for the variance of $\hat{b}$ is exactly equal to 1/4 of the corresponding formula for the CAR process (i.e. with $\hat{b}$ instead of $\hat{c}$ in the formula above).