Spatio-Temporal Models on the Basis of Innovation Processes
and Application to Cancer Mortality Data

Ulrike Schach
Department of Statistics, University of Dortmund, 44221 Dortmund, Germany
uschach@statistik.uni-dortmund.de

Abstract

The aim of this paper is to find a modeling approach for spatially and temporally structured data. The spatial distribution is considered to form an irregular lattice with a specified definition of neighborhood. Additional to the spatial component, a temporal autoregressive parameter, and a time trend are modeled within a multivariates Markov process. This Markov process can be expressed on the basis of an innovation process, which allows for statistical inference on various parameters.

Keywords: Lattice data, conditional autoregressive approach, spatio-temporal linear model, innovation process, ML-estimation

1 Introduction and Structure of the Data

Modeling phenomena dependent on space and time can be done in several different manners. The challenge is to combine time series theory in a sensible way with the analysis of spatial structures and suitable covariates. The distribution in space and time needs to be modeled simultaneously, in order to gain knowledge about the spatial and temporal parameters. The data that has guided this investigation are stomach cancer mortality data among men for the Federal Republic of Germany (west), provided by the German Cancer Institute in Heidelberg. They are counts data collected spatially on the basis of administrative units called "Regierungsbezirke", of which there are 30 in former West Germany and available for a 15-year time period from 1976 to 1990 on a yearly basis. Additional to the response variable there are several possible variables of influence available. As stomach cancer is mainly influenced by nutrition and living conditions [1], p. 51, and since these variables are difficult to obtain, the population density will be taken as a surrogate. According to Kafadar and Tukey [7], the population density can be considered as an indication of a region's level of urbanization. They showed that urban and rural areas differ with respect to various types of cancer, and they suggest to take the logarithm of the population density in order to adjust the scale of the covariate to the one of the mortality rates.

1.1 Standardization

The pure stomach cancer mortality counts of every region need to be standardized with respect to age group and gender, as the occurance of death due to cancer depends on these variables to a great extent. A standardization guarantees the comparability between regions and years, by assuming similar population structures within the study populations. For the analysis of the underlying data set, a mixture between internal and external standardization has been chosen in a way that the standard population has been calculated by summing up the population numbers over 15 years from 1976 to 1990 for age group and gender. The advantage of this kind of standardization is that the temporal trend within the data can be conserved. A standardization of the regional counts using the yearly population automatically leads to a removal of that trend. Additionally the data will be standardized indirectly, see Kreienbrock & Schach [9], p.36 ff. Therefore, consider the following notation, where k denotes age group, with $k = 1, \ldots, K$, i is the spatial index and runs from $i = 1, \ldots, D$ through the 30 regions of Germany, and $t = 1, \ldots, T$ is the temporal index.

M_{kit} := number of deaths of the defined cancer in the study population
N_{kit} := number of people in the study population
M^*_{kit} := number of deaths of the defined cancer in the standard population
N^*_{kit} := number of people in the standard population

As it is the aim to calculate the standardized mortality ratio (SMR), an indirect method of standardization needs to be used. The SMR can be interpreted as a natural ratio of observed cases divided by expected cases, as the SMR can be expressed as

$$\mathrm{SMR}_{it} = \frac{\mathrm{MR}_{it}}{\mathrm{MR}_{\mathrm{ind}_{it}}},$$

with

$$\mathrm{MR}_{it} = \frac{\sum_{k=1}^{K} N_{kit}\, \mathrm{MR}_{kit}}{\sum_{k=1}^{K} N_{kit}}. \tag{1}$$

MR_{kit} is calculated as the quotient of M_{kit} divided by N_{kit}. Then (1) can be simplified to

$$\mathrm{MR}_{it} = \sum_{k=1}^{K} W_{kit}\, \mathrm{MR}_{kit}.$$

The indirectly standardized rates are obtained as follows

$$\mathrm{MR}_{\mathrm{ind}_{it}} = \sum_{k=1}^{K} W_{kit}\, \mathrm{MR}^*_{kit} \qquad \text{with } W_{kit} = \frac{N_{kit}}{\sum_{k=1}^{K} N_{kit}}. \tag{2}$$

Again, the mortality rate MR^*_{kit} of the standard population is given by the quotient $\frac{M^*_{kit}}{N^*_{kit}}$. Since the SMR's show a strong linear dependence of the standard deviation on the mean, a logarithmic transformation has been chosen. This transformation leads to a reduction of this dependence, so that theoretical model assumptions have more validity. Later on, the overall mean will be subtracted from the data, as an expected value of 0 is required for the modeling approaches.

Figure 1 shows the temporal trend of the transformed standardized stomach cancer mortality data. Each box stands for the summarized data of all sites within one year. Clearly, one can see the downward trend. For further analyses, this trend needs to be estimated and removed since especially the small scale variation within the data is of interest. In figure 2, the logged SMR's of the 30 sites are displayed. Here, the data have been aggregated over 15 years in order to build the boxes. Obviously there is still a considerable amount of variation between the sites. Especially the sites with the numbers 24 to 29 show relatively high rates. These sites are all located in the south of Germany, and according to cancer specialists it is not well understood what causes this behaviour.

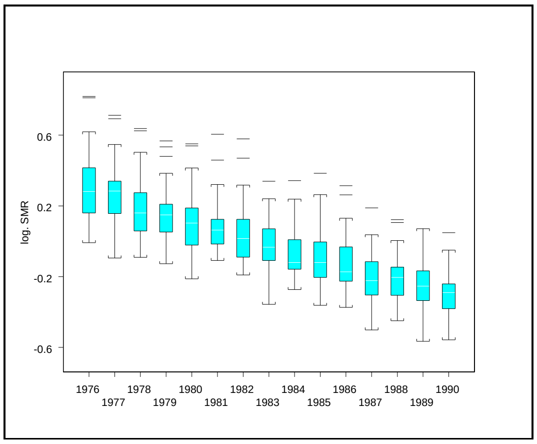

Figure 1: Temporal distribution of logged stomach cancer SMR's of men aggregated over the 30 regions of Germany

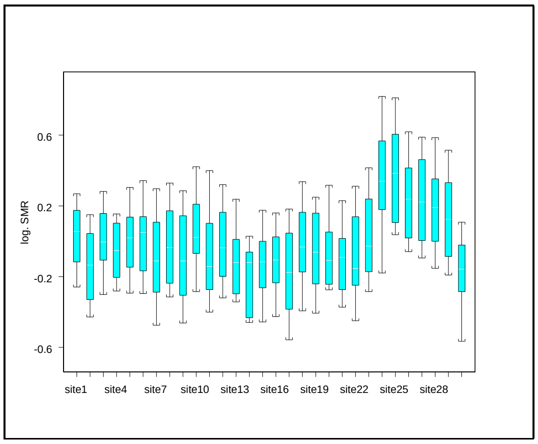

Figure 2: Spatial distribution of logged stomach cancer SMR's of men aggregated over the 15 year time period

1.2 Lattice Data

The data underlying this analysis are so called lattice data as they form an (irregular) lattice in $I\!R^2$, where the data of a certain region within that lattice consists of the logged stomach cancer SMR's. Figure 3 shows the spatial structure for the three years of 1976, 1983 and 1990.

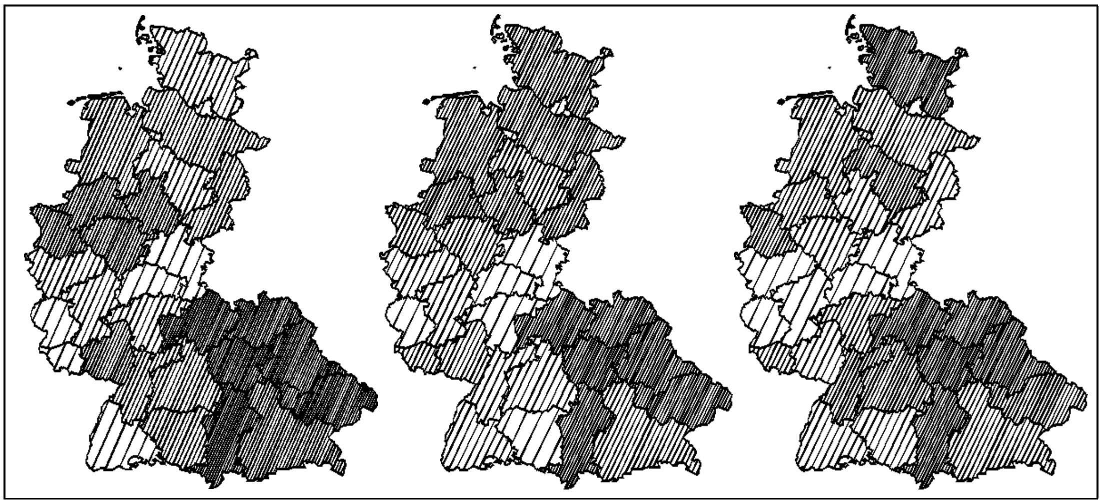

Figure 3: Spatial structure of stomach cancer mortality rates for the years 1976, 1983 and 1990

The most important feature of lattice data is the definition of neighborhood structures. Two regions are considered to be neighbors, if they share a common border. A different approach can be obtained from so called geostatistical data, where the *location* of the data and the distance between them is of special interest. A neighborhood structure can be defined which considers two locations as neighbors, if they lie within a certain distance. If this theory is transferred to lattice data, the center of every region, represented either by the "gravity" center or the main city of it, is considered to contain all the information about the region. Thus, two regions are neighbors, if their centers lie within a certain distance of each other. For further information on geostatistical theory, see Markus et al. [10]. A typical feature of geographical data, collected in adjacent regions, is the dependence of the observations, i.e. the mortality rates in this case. The dependence of the data can either be caused by similar environmental conditions in neighboring areas or through a real influence of one area on its neighbors. So the definition whether two sites are neighbors is particularly important for (spatially) dependent data, as it allows to account for the dependence structure, i.e. model it.

1.3 Conditional Modeling Approaches

Let $\{Z_i : i \in D\}$ be a pure spatial process. Such processes are analogues of time series models [3], therefore time series theory can be applied to some extent. Especially, transformation of a notion of the Markov dependence is of great importance for spatial processes. In time series, a random process has the Markov property if its future observations given the observations at present do not depend on the observations in the past. With respect to the spatial lattice and a definition of neighborhood, the spatial Markov property can be defined as follows: the outcome of region i given the outcome of its adjacent regions does not depend on the outcomes of all the non contiguous regions of the lattice. A model based on this assumption is called conditionally autoregressive (CAR). A different approach, which is not persued here, uses the simultaneous distribution of the data on the lattice, as described by Besag [2].

However, the data does not only have a spatial but also a temporal structure, the cancer mortality rates for every site are available for a period of 15 years from 1976 to 1990. Therefore the underlying stochastic system can either be considered as a family of spatial distributions $\{\{Z_{i,t} : i \in D\}, t \in T\}$ with a temporal index $t = 1, \ldots, T$. Or it can be expressed through a family of time series $\{\{Z_{t,i} : t \in T\}, i \in D\}$ with a spatial index $i = 1, \ldots, N$, see Pfeiffer & Deutsch [11].

2 Spatio-Temporal Gaussian Models

2.1 Spatial and Temporal Dependence

The aim is to model data dependent on space and time by using the theory of stochastic processes. The simplest model separates spatial and temporal effects additively and can be written as

$$Z_t = \alpha\, Z_{t-1} + \beta\, B\, Z_{t-1} + \epsilon_t, \qquad t = 1, \ldots, T. \tag{3}$$

α and β are the parameters of the temporal respectively the spatial autocorrelation of order one. B is the neighborhood matrix, i.e. $B \sim (n, n)$ is of the following form: the (i, j)th element of B is 0, if site i and site j are not neighbors, or if $i = j$. Else, if i and j are neighbors, the (i, j)th element of B is $\frac{1}{n_i}$, where n_i is the number of neighbors of site i. This kind of "weighting" is necessary to ensure that every site is influenced by its neighbors to the same extent. $\epsilon_1, \epsilon_2, \ldots$ are an iid. innovation sequence, and especially ϵ_t is independent of $Z_1, \ldots, Z_t$. Using $C = C_{\alpha\beta} = \alpha\, I + \beta\, B$ the process in (3) can be rewritten as

$$Z_t = C\, Z_{t-1} + \epsilon_t \qquad t = 1, \ldots, T, \tag{4}$$

where C contains both the spatial and the temporal information.

2.2 Representation of Z_t on the Basis of an Innovation Process

Let Z_t be a multivariate stochastic process with the following characteristics:

i) $\{Z_1, \ldots, Z_T\} \sim \text{ Gau}(0, \tilde{\Sigma})$

ii) $\{Z_t\}$ has the Markov property

iii) $\{Z_t\}$ is second order stationary.

Define $\Sigma := \text{cov}(Z_t)$. This is possible, since the covariance matrices do not depend on t, due to the second order stationarity of the process Z_t. It follows from assumption i) that $E(Z_t) = 0$. Additionally define $\Delta := \text{cov}(Z_t, Z_{t-1}) = E(Z_t\, Z'_{t-1})$. According to Fahrmeir [4], p. 27 , it can be shown that

$$E(Z_t \mid Z_{t-1}) = \Delta\, \Sigma^{-1}\, Z_{t-1} \tag{5}$$

and

$$\begin{aligned}
\text{cov}(Z_t \mid Z_{t-1}) &= \text{cov}(Z_t) - \text{cov}(Z_t, Z_{t-1})\, \text{cov}(Z_{t-1})^{-1}\, \text{cov}(Z_{t-1}, Z_t) \\
&= \Sigma - \Delta\Sigma^{-1}\Delta'.
\end{aligned} \tag{6}$$

Let a sequence of multivariate normal random vectors ξ_t, $-\infty < t < \infty$, be given with $E(\xi_t) = 0$ and $\text{cov}(\xi_t) = \Sigma - \Delta\, \Sigma^{-1}\, \Delta' =: \Sigma_0$, where $\xi_0, \xi_1, \ldots, \xi_t$ are independently and identically distributed. Define $\tilde{Z}_t$ as

$$\tilde{Z}_t = \sum_{j=0}^{\infty} (\Delta\, \Sigma^{-1})^j\, \xi_{t-j}.$$

Obviously $E(\tilde{Z}_t) = \sum_{j=0}^{\infty} (\Delta\, \Sigma^{-1})^j\, E(\xi_{t-j}) = 0$ and

$$\begin{aligned}
\text{cov}(\tilde{Z}_t) &= \sum_{j=0}^{\infty} (\Delta\, \Sigma^{-1})^j\, (\Sigma - \Delta\, \Sigma^{-1}\, \Delta')\, (\Delta\, \Sigma^{-1})'^j \\
&= \sum_{j=0}^{\infty} (\Delta\, \Sigma^{-1})^j\, \Sigma\, (\Sigma^{-1}\, \Delta')^j - \sum_{j=0}^{\infty} (\Delta\, \Sigma^{-1})^{j+1}\, \Sigma\, (\Sigma^{-1}\, \Delta')^{j+1} \\
&= \Sigma.
\end{aligned}$$

Since $\tilde{Z}_t$ has the same moments as Z_t, and since Z_t and $\tilde{Z}_t$ are normally distributed, the two processes are identical with $C_{\alpha\beta} = \Delta\,\Sigma^{-1}$. Thus Z_t can be written as

$$Z_t = \Delta\,\Sigma^{-1}\,Z_{t-1} + \xi_t.$$

It becomes clear now that Z_t can be expressed recursively through the constant matrix C containing the spatial and temporal structure of the process, based on the initial distribution at time $t = 0$, plus the innovation term ξ_t, $t = 1, \ldots, T$. However, there need to be specified some conditions for the parameters α and β, to guarantee that the process Z_t converges. It follows from $C_{\alpha\beta} = \Delta\,\Sigma^{-1}$ that $\Sigma_0 = \Sigma - C\,\Sigma\,C'$. Then Σ can be written as follows

$$\Sigma = \mathrm{cov}(Z_t) = \sum_{j=0}^{\infty} C^j\,\Sigma_0\,(C')^j. \tag{7}$$

A sufficient condition for convergence of $\mathrm{cov}(Z_t)$ is that a suitable matrix norm of C is smaller than 1. This criterion is dependent on the parameters α and β, as they determine C. The spectral matrix norm, with

$$\|D\| := \max\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } D'D\}$$

has been chosen for this problem. According to Horn & Johnson [6], p. 295f., the spectral norm satisfies the triangle inequality and hence

$$\begin{aligned} \|C\| = \|\alpha\,I + \beta\,B\| &\leq \|\alpha\,I\| + \|\beta\,B\| \\ &\leq \mid \alpha \mid\ \|I\| + \mid \beta \mid\ \|B\|. \end{aligned}$$

The spectral norm of matrix $B'B$ is 1.089. The spectral norm of the identity matrix is 1. Therefore, the following sufficient condition for convergence of Σ on the parameters α and β can be given by

$$\mid \alpha \mid + 1.089 \mid \beta \mid < 1.$$

In that case, convergence of Σ follows from the convergence of the geometric series as is shown below:

$$\begin{aligned} \|\sum_{j=m}^{n} C^j\,\Sigma_0\,(C')^j\| &\leq \sum_{j=m}^{n} \|C^j\,\Sigma_0\,(C')^j\| \\ &\leq \sum_{j=m}^{n} \|C^j\|\,\|\Sigma_0\|\,\|(C')^j\| \\ &= \|\Sigma_0\| \sum_{j=m}^{n} \|C\|^j\,\|C'\|^j \to 0, \qquad \text{for } m, n \to \infty. \end{aligned}$$

3 Data Analysis

Considering the underlying model for Z_t, given by assumptions i) to iii)

$$Z_t = \alpha\, Z_{t-1} + \beta\, B\, Z_{t-1} + \epsilon_t$$

of interest are the parameters α and β and matrix Σ. Taking the underlying data set into consideration, assumptions i) to iii) need to be examined. The normality assumption of i) can be justified for the following reason: a weighted sum of independent random variables is asymptotically normal under suitable regularity conditions by the central limit theorem. In formulas (1) and (2) quantities arising from the standard population are considered to be constant. Only the mortality counts of the age groups are random variables; and the SMR's can easily be seen to consist of weighted sums of these variables. Hence approximate normality follows. The expected value of this normal distribution can be taken as 0, since the overall mean has been subtracted from the data. The Markov property, see ii), has been assumed since the partial autocorrelation of the data indicates a lack of dependence of future and past given the present. Figure 4 shows the autoregressive structure of order one for three selected German regions using partial autocorrelation functions.

Finally, the process can be considered to be stationary, because stomach cancer mortality rates have existed a long time before the year of 1976, when the time series of this analysis starts and the process has been documented. Therefore, the process is in its equilibrum already. As $\{Z_t\}$ is second order stationary and Gaussian, it follows that Z_t is strongly stationary. $Z_1, \ldots, Z_T$ are identically distributed and the covariance between two points in time only depends on their distance.

3.1 Likelihood Function

Due to the Markov property of the stochastic process Z_t, the likelihood function can be written as a product of the conditional transition probabilities, although they are not independent. A so far unknown matrix is Σ, the covariance matrix of Z_t. This, however, does not turn out to be a problem, since Σ can be expressed through matrix C and a simplification of the covariance of the innovation Σ_0. It is assumed from here, that $\Sigma_0 = \sigma_0^2\, I$. Then Σ can be expressed by

$$\begin{aligned} \Sigma = \mathrm{cov}(Z_t) &= \sum_{j=0}^{\infty} C^j\, \Sigma_0\, (C')^j \\ &= \sum_{j=0}^{\infty} C^j\, \sigma_0^2\, I\, (C')^j \\ &= \sigma_0^2 \sum_{j=0}^{\infty} C^j\, (C')^j. \end{aligned}$$

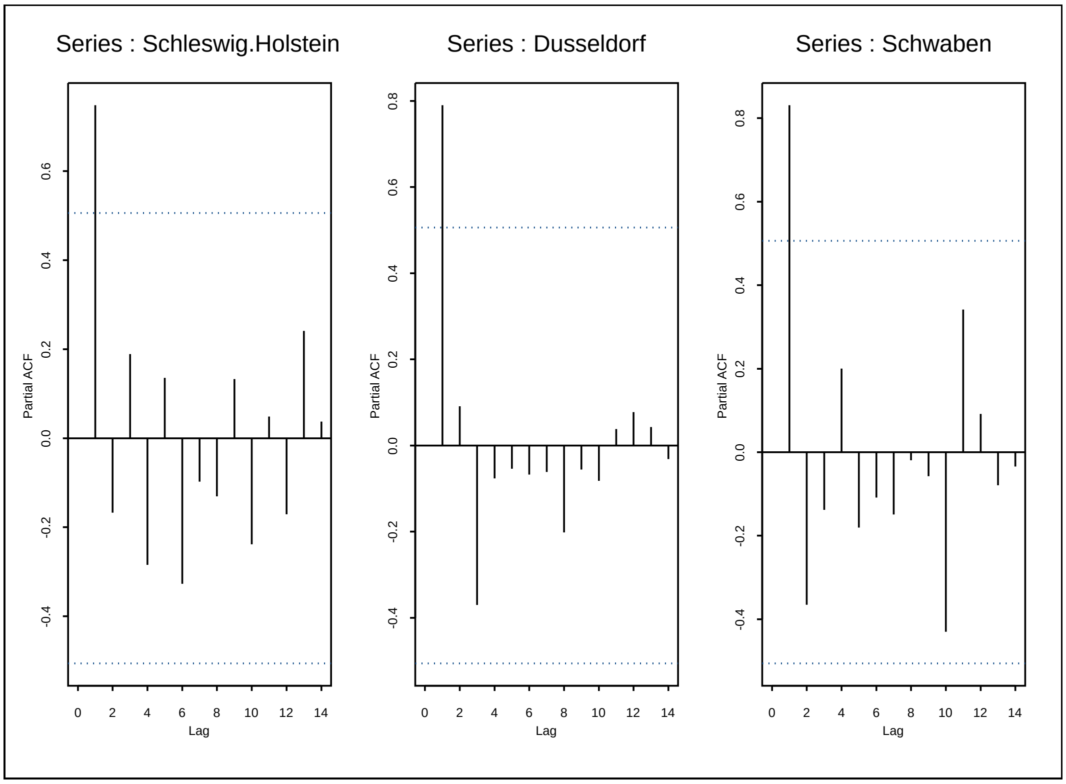

Figure 4: Partial temporal autocorrelation for the regions Schleswig-Holstein, Düsseldorf and Schwaben

As a justification for this, ϵ_t can be considered as noise, perhaps due to measurement and observation errors, which acts on the components of the process. It is reasonable to assume, that all components have the same variance and are independent of each other. Thus, with starting values for α, β, and σ_0^2, and a reasonable number of replications (terminated by a stopping rule dependent on the spectral norm of matrix C), a matrix Σ will be obtained, that can be considered as a close approximation to the covariance matrix of the process at time t.

Based on the modification of the covariance matrix, the likelihood function can be expressed as

$$\begin{aligned} l(\alpha, \beta, \sigma_0^2 \mid Z_1, \ldots, Z_T) \quad &= \quad \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \Sigma}} \exp\{-\frac{1}{2} Z_1' \, \Sigma^{-1} \, Z_1\} \\ & \quad \prod_{t=2}^{T} \frac{1}{(2\pi)^{\frac{N}{2}} \, (\sigma_0^2)^{\frac{N}{2}}} \exp\{-\frac{1}{2} \frac{1}{\sigma_0^2} \, \|Z_t - C \, Z_{t-1}\|^2\}. \end{aligned}$$

3.2 Generalization of the Likelihood to the Case of Covariates

When including a matrix of covariates and a vector $\gamma = (\gamma_0, \ldots, \gamma_p)'$ of regression coefficients in the model and the likelihood function belonging to it, two different cases can be formulated. In the first case, the covariates are constant over the observed period of time. So the likelihood function can be generalized to

$$\begin{aligned} l(\alpha, \beta, \gamma, \sigma_0^2 \mid Z_1, \ldots, Z_T) \quad = \quad & \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \Sigma}} \exp\{-\frac{1}{2} (Z_1 - X\,\gamma)' \, \Sigma^{-1} \, (Z_1 - X\,\gamma)\} \\ & \prod_{t=2}^{T} \frac{1}{(2\pi)^{\frac{N}{2}} \, (\sigma_0^2)^{\frac{N}{2}}} \exp\{-\frac{1}{2} \frac{1}{\sigma_0^2} \|Z_t - C\, Z_{t-1} - X\,\gamma\|^2\}, \end{aligned}$$

where $\gamma_0, \ldots, \gamma_p$ are the unknown regression coefficients and X is the *constant* regressor matrix. In the second case, even the covariates have a temporal structure. They can be considered to form a separate stochastic process over the observed time period. In that case, the likelihood will be written as

$$\begin{aligned} l(\alpha, \beta, \gamma, \sigma_0^2 \mid Z_1, \ldots, Z_T) \quad = \quad & \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det \Sigma}} \exp\{-\frac{1}{2} (Z_1 - X^{(1)}\,\gamma)' \, \Sigma^{-1} \, (Z_1 - X^{(1)}\,\gamma)\} \\ & \prod_{t=2}^{T} \frac{1}{(2\pi)^{\frac{N}{2}} \, (\sigma_0^2)^{\frac{N}{2}}} \exp\{-\frac{1}{2} \frac{1}{\sigma_0^2} \|Z_t - C\, Z_{t-1} - X^{(t)}\,\gamma\|^2\}. \end{aligned}$$

As described in section 1, the underlying covariates are constant in this analysis and therefore the likelihood of the first case will be used for future analyses. Additionally, the temporal trend of the SMR's will be estimated within the model.

3.3 Maximum Likelihood Estimation

The idea of maximum likelihood (ML) estimation is to find those estimators of the unknown parameters $\theta = (\theta_0, \ldots, \theta_k)'$, that have their maximum probability, given the data. The ML method will be used here, because it has good asymptotic behaviour under relatively weak regularity assumptions. It can be shown that the ML estimator is asymptotically normal, consistent and sufficient for the unknown parameters θ, given independent and identically distributed data. Even in this case, where the data underlying the process are indentically distributed, but not independent, as described in section 3.1, the ML estimators are optimal. Using Martingale limit theory it can be shown that the ML estimators are asymptotically normal, consistent and sufficient estimators, see Hall & Heyde, [5], p. 156. So the aim is to find a local or global maximum of the likelihood function, or the logarithm of the likelihood function for a faster calculation. A necessary condition for that is

$$\frac{\partial l}{\partial \theta}(\hat{\theta}; Z) = 0.$$

However, it is not always possible to solve the above equation analytically, especially considering a multivariate problem. Therefore the estimates sometimes have to be calculated using a numerical or iterative procedure as the quasi-Newton method, which is used by S-Plus in order to avoid calculating the Hessian matrix of second derivatives. As the quasi-Newton method is endangered to provide local instead of global extrema, different starting values will be given, to obtain reliable results.

3.4 Tests and Confidence Intervals

Having calculated the ML estimates for the unknown parameters, it is of interest, to test for their significance and build confidence intervals. Therefore, consider the Hessian at the estimated points. It contains the estimated variances of the estimated parameters on the diagonal and their estimated covariances off diagonal. As stated above, it can be shown that the ML estimators are asymptotically normal. This will be used for the construction of confidence intervals and statistical tests.

4 Application to the Data

The application of the described model to the data has been done for the case without covariates. The results of the parameter estimation for α, β, σ_0 and the temporal trend are displayed in table 1.

Parameters estimates			
Temporal AC α	Spatial AC β	σ_0	Temporal trend
0.7029	0.2915	0.0931	-0.0041

Table 1: Results of the parameter estimation without covariates

The estimated parameters of table 1 seem to be global maxima as they are reproducible, independently of their different starting values. With an average number of 40 iterations, the estimated parameters are identical up to the 7th decimal number. With a considerable amount of calculation, it is possible to find the estimated variance-covariance matrix of these parameters. Especially, if the Hessian is not supplied, the variance-covariance matrix is difficult to obtain. Table 2 gives the values of the test statistic for each parameter.

Values of the test statistics			
Temporal AC α	Spatial AC β	σ_0	Temporal trend
17.19	6.91	29.71	-8.39

Table 2: Test statistic values for the estimated parameters

Obviously, all four parameters are significant at a 5 percent level and confidence intervals can be given as follows.

α	:	(0.7029	$\pm$	0.080)
β	:	(0.2915	$\pm$	0.083)
σ_0	:	(0.0931	$\pm$	0.006)
temp. trend	:	(-0.0041	$\pm$	0.001)

Table 3: 95% confidence intervals for the estimated parameters

5 Discussion

The examination of male stomach cancer mortality rates in Germany over the 15 year time period from 1976 to 1990 shows a relatively strong temporal autocorrelation of 0.7029. The estimated spatial autocorrelation with 0.2915 is much smaller. The estimation of the regression coefficients (for an overall trend, the logged population densitiy, the unemployment rate and the logged gross domestic product) has lead to enormous numerical problems, which have so far been unable to solve. Especially when trying to maximize the Likelihood function using the quasi-Newton method, convergence is not easy to obtain. Therefore, the parameter estimate for β must be viewed with caution. Parts of the spatial autocorrelation β will be caused by underlying covariables with a spatial correlation structure, that are yet unaccounted. The estimation of the temporal trend is presumably not affected by missing covariates, as the covariates are considered to be constant over the 15 year time period. Future research should be undertaken in the direction of covariates with a temporal trend, if they can be obtained. Furthermore, the temporal and spatial autoregressive structure of the described model is of order one. Several other approaches are imaginable, such as to include a dependence of higher order, or to model the spatial autocorrelation with no temporal lag.

Different models for the same problems can be found with a Bayesian point of view. Especially by simulating from the a posteriori distribution using MC-methods instead of calculating it, computation times can be shortened, see Krause [8], p.1 ff., which is of great advantage compared to classical modeling.

Acknowledgements

This research was supported by the German Research Council (DFG) through the Graduate College and the Collaborative Research Centre at the University of Dortmund (SFB 475): Reduction of complexity for multivariate data structures.

References

[1] Becker, N. (1984): *Atlas of Cancer Mortality in the Federal Republic of Germany.* 2nd edition. Springer-Verlag, Berlin

[2] Besag, J. (1974): Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Association B, 36*, 192-225.

[3] Cressie, N. (1993): *Statistics for Spatial Data.* Wiley, New York

[4] Fahrmeir, L. & Hamerle, A. (1984): *Multivariate statistische Verfahren.* de Gruyter, Berlin

[5] Hall, P. & Heyde, C. (1980): *Martingale Limit Theory and Its Application.* Academic Press, New York

[6] Horn, R. & Johnson, C. (1985): *Matrix Analysis.* Cambridge University Press, Cambridge

[7] Kafadar, K. & Tukey, J. (1993): U.S. Cancer Death Rates: A Simple Adjustment for Urbanization. *International Statistical Review 61*, 259-281.

[8] Krause, A. (1994): *Computerintensive statistische Methoden.* Gustav Fischer Verlag, Stuttgart

[9] Kreienbrock, L. & Schach, S. (1994): *Epidemiologische Methoden.* Gustav Fischer Verlag, Stuttgart

[10] Márkus, L.; Berke, O.; Kovács, J. & Urfer, W. (1999): Spatial Prediction of the Intensity of Latent Effects Governing Hydrogeological Phenomena. *Environmetrics 10*, 633-654.

[11] Pfeiffer, P. & Deutsch, S. (1980): Identification and Interpretation of First-Order Space-Time ARMA Models. *Technometrics 22*, 397-408.