Raport Badawczy
Research Report

RB/18/2008

A statistical model for spatial inventory data: a case study of N_2O emissions in municipalities of southern Norway

J. Horabik, Z. Nahorski

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2008

A statistical model for spatial inventory data: a case study of N_2O emissions in municipalities of southern Norway

Joanna Horabik, Zbigniew Nahorski
Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
Tel: 4822 3810 100, Fax: 4822 3810 105
{Joanna.Horabik, Zbigniew.Nahorski}@ibspan.waw.pl

Abstract. In this paper we apply a linear regression with spatial random effect to model spatially distributed emission inventory data. The topic is related to the issue of disaggregation of national greenhouse gas emissions into fine spatial grid. Emission maps are typically produced from information on spatially explicit activities contributing to emissions, which are multiplied by emission factors. In our case study we have available N_2O emission assessments for municipalities of southern Norway, as well as three kinds of covariate information for each region. Thus a regression model can be fitted to these data. We use conditionally autoregressive model to account for spatial correlation between municipalities. Estimation of parameters is based on the Bayes Theorem and the Gibbs sampler algorithm. The results suggest that one of initially considered covariates should be excluded from further analysis, and instead presence of another, spatially correlated factor(s) is suggested. Moreover, the model was capable to capture an outlier point source emission from nitric acid plant. The point of the contribution is that including spacial information helps in finding right explanatory variables. This way improved emission assessments can be provided based on future covariate data.

1. Introduction

Our contribution is focussed on a spatial aspect of inventories for atmospheric pollutants. The study is motivated with situations when two kinds of inventories for the same area and for the same pollutant are available: (i) based on detailed information on emissions (sometimes referred to as a bottom-up approach) and (ii) based on spatial activity information (e.g. land use, traffic etc.) multiplied by appropriate emission factors. Provided both data sets were obtained independently, one may compare two maps and try to conclude on the relevance of activity data used. To this end we take a statistical approach.

This kind of analysis has been already performed in some studies. Specifically, we were motivated with the publication of (Winiwarter et al., 2003). In that article two sets of data on NOx (Nitrogen oxides) emissions over the same spatial grid for the Greater Athens, Greece

were compared. The authors examine significance of area, line and point emission sources on the basis of statistical exploratory tools and a visual comparison of maps.

In the case of greenhouse gasses, spatial resolution is not crucial for the emission effect as such. However, there are several situations when spatial dimension is needed. In elaborated models of climate change, like British Meteorological Office model HadAM3 (Pope et al., 2000), transport of greenhouse gasses is modelled similarly to other pollutants, like SO_x and NO_x. With growing resolution, for instance in national models of this type, the need for finer mesh of inventory data becomes important. The proposed method can be used for this purpose. Other examples include validations of regional inventories by field measurements or by inverse modelling of satellite data.

The topic of spatial heterogeneity of greenhouse gas emissions and sequestration can be addressed in various ways. For instance, the spatial distribution of greenhouse gas emissions for Ukraine has been presented in (Bun et al., 2007). (Theloke et al., 2007) develop a methodology for spatial (and temporal) disaggregation of greenhouse gas annual country totals. (Oijen and Thomson, 2007) consider using a process-based forest model which accounts for spatial distribution of climate and soil; a Bayesian calibration is employed to quantify uncertainties.

When performing a statistical inference of spatial inventory data, we account for the fact that values at proximate locations tend to be more alike. This motivates use of spatial statistics. Moreover, since for each grid cell we have information on aggregated emission values, these are areal data (also known as lattice data). A popular tool for incorporating this kind of spatial information is conditionally autoregressive (CAR) model proposed by (Besag, 1974). As opposed to geostatistical models with spatially continuous data, CAR models have been developed to account for the situation where the set of all possible spatial locations is countable. The idea is to define a model in terms of the conditional distribution of the observation at one location given its values at other neighbouring locations. Applications of CAR model include among others mapping diseases in counties and modelling particular matter air pollution in space and time (Kaiser et al., 2002).

The aim of the present paper is to demonstrate the usefulness of CAR model to analyse data from spatially distributed emission inventory. With available data on factors contributing to emissions and an independent set of emission assessments one may build a suitable regression model. Inclusion of a spatial component is intended to help in finding right explanatory variables, and thus provide more robust emission assessments in situations when only covariate information is available. The outline of the study is the following. Section 2 presents

our illustrative data set and an initial non-spatial model is suggested. Next, it is enriched with a spatial random effect. We use a conditionally autoregressive structure to account for spatial correlation between neighbouring areas (municipalities, in our case). The model is described in Section 3. Results are presented in Section 4 - we fit the spatial model and compare various combinations of covariates. Section 5 contains final remarks.

2. Preliminary explorations

Our illustration is provided with data set on N_2O (nitrous oxide) emissions (in tonnes) reported in municipalities (counties) of southern Norway. The data come from StatBank (available at `http://www.ssb.no`) in Statistics Norway. The map covers 259 municipalities. For each municipality three kinds of covariate information are available. We assume that the N_2O emissions, denoted E, may depend on: population P (number of residents in a municipality), area covered by roads R (in km^2) and the total area of a municipality A (in km^2).

Our original model on municipality emissions has the form

$$E_i = cA_i^{\beta_1} P_i^{\beta_2} R_i^{\beta_3} \tag{1}$$

where c and $\beta_j, j = 1, 2, 3$ are coefficients to be estimated and $i = 1, \ldots, 259$ indexes municipalities.

We use a log-transformation on the emission data and covariates. The log-transformation of the emission data and the covariates enables simpler estimation approach - the linear regression can be used. Let us denote:

- y_i - logarithm of N_2O emissions E_i, $y = (y_1, \ldots, y_n)'$
- $x_{i,1}$ - logarithm of total area A_i, $x_1 = (x_{1,1}, \ldots, x_{n,1})'$
- $x_{i,2}$ - logarithm of population P_i, $x_2 = (x_{1,2}, \ldots, x_{n,2})'$
- $x_{i,3}$ - logarithm of area covered by roads R_i, $x_3 = (x_{1,3}, \ldots, x_{n,3})'$.

Figure 1 shows histograms of these log-transformed data and Figure 2 presents the respective scatterplot matrix. We notice that the relationship between y and x_2 as well as between y and x_3 resembles the linear one. This is not the case for the (log) total area x_1. To begin with, we check a multiple regression with all three covariates:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \varepsilon_i. \tag{2}$$

where ε_i are independent random variables following normal distribution with mean equal 0 and variance σ^2. We distinguish between an

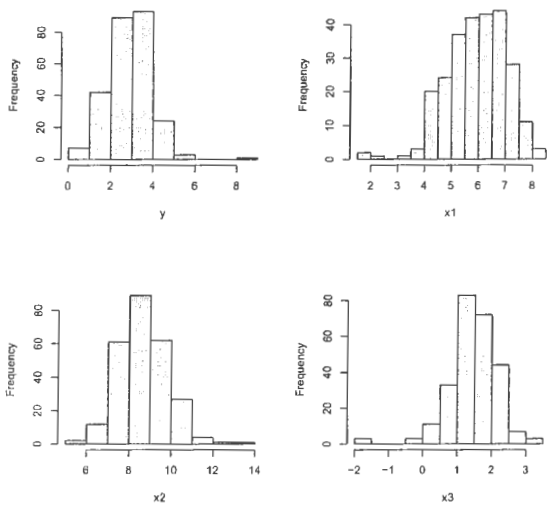

Figure 1. Histograms of log-transformed data on: N_2O emissions (y), total area (x_1), population (x_2) and roads (x_3) in municipalities.

observation (y_i) and a random variable (Y_i) generating this observation. Note that $e^{\beta_0} = c$ and the formula arises from log-transformation of (1).

As to be expected, estimation results, see Table I, showed that the covariate on municipality area x_1 is not significant. Regression coefficients of the remaining two covariates x_2 and x_3 have p-value equal 0.04 and 9.82e-08, respectively. The model explains 47% of variability in log(emissions) - coefficient of determination is $R^2 = 0.47$. Thus, we drop the insignificant variable x_1. We obtain the same coefficient $R^2 = 0.47$ and now both regression coefficients β_2 and β_3 are highly significant (p-value equal 3.14e-08 and 2.53e-15, respectively). Coefficient of determination R^2 for other regression models (i.e. with other combinations of covariates) are lower, which supports the choice of the model with explanatory variables x_2 and x_3 as the best one - at least on this stage of analysis.

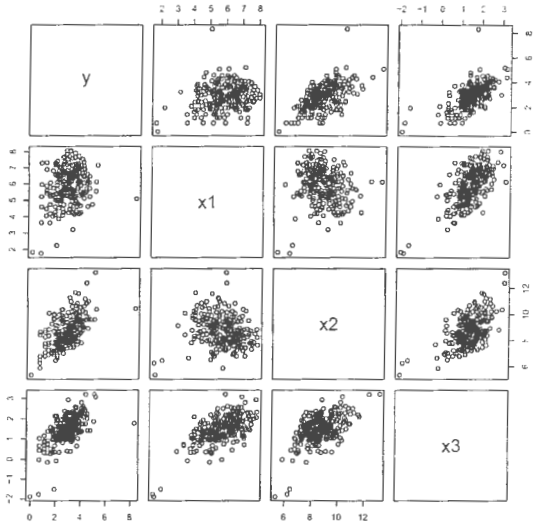

Figure 2. Scatterplot matrix showing plausible relations between log-transformed data on: N_2O emissions (y), total area (x_1), population (x_2) and roads (x_3) in municipalities.

Residuals from this model are presented in Figure 3: a residual plot (a) and a map (b). Two issues should be noted here. First, there is one outlier observation y_{125}, which is the Porsgrunn municipality. We find out that this is due to a nitric acid plant, operating there since 1991, see also (Perez-Ramirez, 2007). Porsgrunn is a relatively small municipality located near the southern coast and on the residual map we spot a striking value. Secondly, from the residual map we can identify clusters of counties with underestimated emissions (yielding positive residuals), and a cluster of counties with overestimated emissions (yielding negative residuals) in the central region.

We check spatial correlation in residuals using Moran's I statistic:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(\varepsilon_i - \bar{\varepsilon})(\varepsilon_j - \bar{\varepsilon})}{\sum_i (\varepsilon_i - \bar{\varepsilon})^2}.$$

a) b)

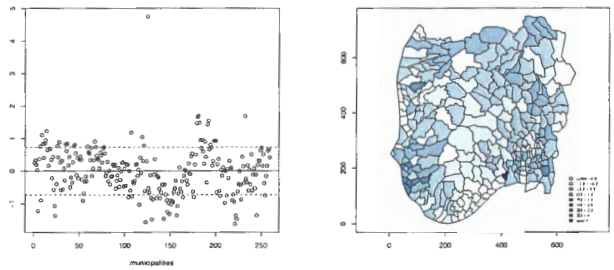

Figure 3. Residuals from the linear model with covariates on (log) population (x_2) and (log) roads (x_3); estimated $\pm\sigma$ displayed as dashed lines.

where ε_i - a residual of linear regression in the area i, $\bar{\varepsilon}$ - the mean of residuals, w_{ij} - the adjacency weights ($w_{ij} = 1$ if j is a neighbour of i, and 0 otherwise, also $w_{ii} = 0$). We consider two municipalities as neighbours if they share common border. Moran's I can be recognized as a modification of the correlation coefficient. It accounts for correlation between residuals in area i and nearby locations and takes values approximately on the interval $[-1, 1]$. Higher (positive) values of I suggest stronger positive spatial association. Under the null hypothesis where ε_i are independent and identically distributed, I is asymptotically normally distributed, with the mean and variance known, see e.g. (Banerjee et al., 2004). In our case of linear regression with covariates on x_2, x_3 Moran's I is equal 0.3144. The corresponding test statistic z (Moran's I standardized with the asymptotic mean and variance) is equal $z = 8.5713$ and $z_{cr} = 2.33$ at the significance level $\alpha = 0.01$. Thus we reject the null hypothesis of no spatial correlation of errors. Moran's I is, however, recommended as an exploratory information on spatial association, rather than a measure of spatial significance (Banerjee et al., 2004).

Table I contains Moran's I statistic calculated for the remaining linear regression models as well. We will discuss it further in the sequel, together with spatial modelling results.

3. Modelling spatial correlation

In this section we develop a Bayesian hierarchical model to characterize spatial distribution of N_2O emissions in municipalities.

Let Y_i denote a stochastic variable associated with the observed logarithm of emission (y_i) defined at each spatial location i for $i = 1, \ldots, n$. The collection of all Y_i is denoted as $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$. It is assumed that the random variables Y_i follow normal distribution with mean μ_i and common variance σ^2. Assumption of normal distribution for (log) emission is equivalent to assuming that emission itself follows lognormal distribution. It is a suitable choice for non-negative continuous processes since lognormal distribution has positive support.

Our approach to modelling the mean μ_i expresses belief that available covariates explain part of the spatial pattern in observations, and the remaining part is captured through a regional clustering. It is expressed as follows. Let the mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ be such that $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\theta}$. Then it is assumed

$$\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\theta}, \sigma^2 \boldsymbol{I}_n\right), \tag{3}$$

where $\boldsymbol{I}_n$ is an identity $n \times n$ matrix. $\boldsymbol{X}$ is the (design) matrix containing a vector of 1s for the intercept β_0 and k explanatory covariates:

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{bmatrix}.$$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ is a vector of regression coefficients. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$ is a vector of correlated random variables. Thus, conditionally on the parameters $\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2$, stochastic variables Y_i are independent.

Next we describe the random spatial component $\boldsymbol{\theta}$. Correlation of variables θ_i allows us to model spatial dependence between the variables Y_i. We make use of the conditionally autoregressive (CAR) model. The CAR structure is given through specification of full conditional distribution functions:

$$\theta_i | \theta_{j, i \neq j}, \tau^2 \sim \mathcal{N}\left(\frac{1}{w_{i+}} \sum_{j \in N_i} \theta_j, \frac{\tau^2}{w_{i+}}\right) \tag{4}$$

with N_i being the set of neighbours of area i, w_{i+} being the number of neighbours of area i and τ^2 is a variance parameter. Conditional expected value of θ_i is the average value of those variables θ_j that are the neighbours of the site i. Conditional variance is inversely proportional

to the number of neighbours w_{i+}. Note that assuming $\theta_i = 0$ for all i would bring us back to the linear regression (nonhierarchical) model considered in the previous section.

Given (4), the joint probability distribution of $\boldsymbol{\theta}$ is the following (Banerjee et al., 2004; Cressie, 1993)

$$\boldsymbol{\theta}|\tau^2 \sim \mathcal{N}\left(\mathbf{0}, \tau^2 (\boldsymbol{D} - \boldsymbol{W})^{-1}\right), \tag{5}$$

where $\boldsymbol{D}$ is $n \times n$ diagonal matrix with w_{i+} on the diagonal and $\boldsymbol{W}$ is $n \times n$ matrix with adjacency weights w_{ij}. Equivalently, after algebraic manipulations (5) can be rewritten as

$$p(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2\tau^2}\sum_{i \neq j} w_{ij}(\theta_i - \theta_j)^2\right]. \tag{6}$$

Estimation of unknown parameters $\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \tau^2$ is based on the Bayes' Theorem. The joint posterior distribution of these parameters is proportional to the product of the likelihood function associated with the formula (3); the distribution of spatial random component $\boldsymbol{\theta}$ in (5); and specified prior distributions for the remaining parameters $\boldsymbol{\beta}$, σ^2 and τ^2. Uniform distributions are employed for each of the $\boldsymbol{\beta}$ parameters. The inverse variance parameters (precisions) $1/\sigma^2$ and $1/\tau^2$ are assumed independent $Gamma(0.01, 0.01)$ distributions, where $Gamma(a, b)$ distribution has the mean equal to a/b. When observations arise from normal distribution, the choice of *Gamma* as a prior for precision leads to posterior precision distribution being also *Gamma*. It is a natural conjugate prior distribution in this case (see e.g. (Press, 2003)).

Combination of all model assumptions allows us to derive full conditional distributions for all parameters in a closed-form, see e.g. (Gamerman and Lopes, 2006). Gibbs sampling is used to update the parameters (Robert and Casella, 2004). Calculations were accomplished both using the `WinBUGS` package (Lunn et al., 2000) and writing our own functions in the statistical software `R` (R Development Core Team, 2008).

4. Results

Spatial CAR models have been applied to the Norway emission data. We estimate spatial models for various combinations of covariates. The results are compared using the Deviance Information Criterion (DIC), which was introduced by (Spiegelhalter et al., 2002). For an intuitive interpretation and a straightforward example see also (Liddle, 2007).

The DIC is calculated as a sum of the posterior expectation of the deviance $\bar{D}$, and the effective number of parameters p_D:

$$DIC = \bar{D} + p_D.$$

$\bar{D}$ measures how well the model fits the data; the larger this value, the worse the fit. Model complexity is summarized by the effective number of parameters p_D - it is a crucially important component of the DIC definition. If all parameters are well-contained within the prior distribution, then p_D is close to the actual number of parameters. Otherwise p_D is smaller than the number of parameters.

The idea of the DIC is to favour a model with a good fit, but also to penalize for the number of parameters. Thus models with smaller DIC are preferred to models with larger DIC.

Table I. Model comparison using coefficient of determination R^2, Moran's I and DIC statistics for linear regressions and spatial CAR models.

Covariates	Linear regression			CAR	
	R^2	DIC	I	p_D	DIC
x_1, x_2, x_3	0.47	571	0.31	190	345
x_1, x_2	0.42	598	0.30	151	426
x_1, x_3	0.46	573	0.32	217	330
x_2, x_3	0.47	572	0.31	199	356
x_3	0.40	601	0.37	227	319
x_2	0.32	633	0.31	132	482

For the estimated spatial CAR models both, the DIC and the effective number of parameters p_D, are displayed in Table I. For the linear regression models p_D is equal to the actual number of unknown parameters (e.g. 5 for a model with 3 covariates: 4 components of β plus variance σ^2). For these models we report only the DIC.

In terms of the DIC criterion, all of the CAR models are much better than any of the initially considered regressions. The hierarchical structure with the spatial random effect performs considerably better. The best result (the lowest DIC equal 319) was obtained for the spatial model only with information on the area covered by roads x_3. This model outperforms among others the CAR model with covariates x_2, x_3. In case of a non-spatial regression these two covariates provided the best results (recall that we rejected the model with covariates x_1, x_2, x_3 since x_1 was not significant). This suggests that some miss-

ing, spatially correlated variables contribute to overall emissions much better than the variable x_2 (log) population.

As already mentioned, Table I shows calculated Moran's I for the considered non-spatial models. The highest $I = 0.37$ is reported for the model with the same set of covariates (just x_3) as the model with the lowest DIC. The applied spatial structure confirmed the exploratory analysis based on the Moran's I tool.

Table II. Parameter estimates for linear regression and respective spatial model (with 95% credible sets).

	Linear regression (x_3)	model CAR (x_3)
β_0	1.51 (1.291, 1.737)	1.44 (1.268, 1.607)
β_3	0.93 (0.801, 1.062)	0.98 (0.871, 1.092)
σ^2	0.59 (0.500, 0.700)	0.06 (0.015, 0.223)
τ^2	-	1.27 (0.810, 1.962)

Another benefit of fitting the spatial structure is seen in the reduction of the length of the 95% credible sets for the covariates in the CAR model compared with the respective non-hierarchical model, see Table II. We show results for the spatial model CAR (x_3), which gave the best results. For the intercept a reduction in length of the 95% credible set is approximately 24%, and for the effect of (log)roads it is around 15%.

Figure 4 shows maps with (a) the posterior mean for the best spatial model i.e. CAR (x_3), (b) fitted values for the best linear regression model with covariates x_2, x_3, and (c) the observations. It can be noticed that the spatial model maps the original data much better. The outlier emission in the Porsgrunn municipality is now captured. Most of the remaining locations with somewhat higher emissions have been correctly identified as well.

Figure 5 depicts the spatial random effect θ_i. It illustrates the pattern of the suggested still-missing explanatory variable(s) and may help in its identification.

5. Concluding remarks

We have shown an application of the spatial conditionally autoregressive structure to examine dependence of covariate data on an independent bottom-up inventory. In the considered case our results suggest

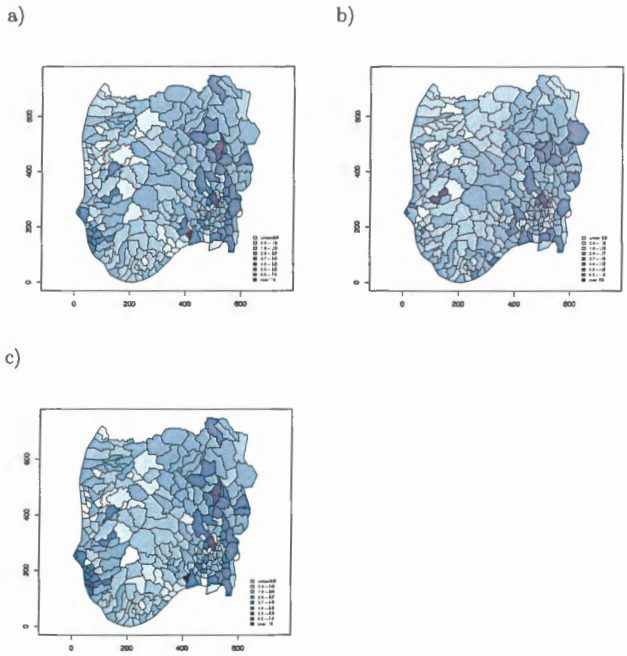

Figure 4. Posterior mean of log(emission) for model CAR with covariate x_3 (a); fitted values in linear regression with covariates x_2, x_3 (b); observed (log) emission (c).

excluding an initially considered covariate on a population of municipalities. Instead, the presence of another missing, spatially correlated factor(s) is suggested. Generally, such situation - that we get better results just for a subset of covariates plus a spatial component - is not unusual. The point of the contribution was to take advantage of the spatial resolution in evaluation of inventory data. Statistical hierarchical models, estimated within the bayesian framework offer a flexible tool.

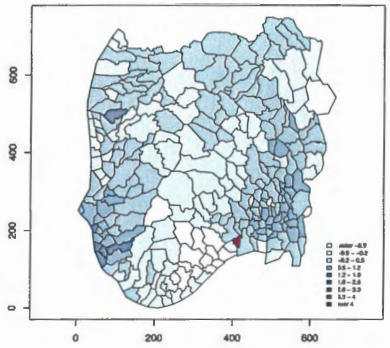

Figure 5. Spatial random effect θ_i, CAR model with covariate on roads (x_3)

This kind of evaluation can help to improve future emission assessments through: (i) indication of the most relevant covariates, and (ii) estimation of the spatial random effect $\boldsymbol{\theta}$. Provided the estimated relationship could last for some time, the model can be used to produce improved emission assessment based on another set of covariates.

A potentially problematic part of inventory are emission point sources (plants), which are correctly reported in a bottom-up approach (as a response) but are missing in datasets with activity information (Winiwarter, 2007). The proposed method proved to be capable to identify such situations, as was the case of the Porsgrunn nitric acid plant.

In the considered model different sizes of the spatial areas (municipalities) were not relevant for the emission level - the covariate on total area of municipality was not statistically significant. In this sense we simply ignore that sizes are different. However, we noticed (not shown) that in some of the models with covariate on area size, the corresponding regression coefficient was negative (on a log scale), which provided a kind of weighting on the emission scale.

It is noted again that in our example the models were applied on the log-transformed response and the covariates. It converts the original multiplicative relationship of the covariates to the linear one. However, this fact does not influence our conclusions on the choice of the factors. Usually emission inventories are carried as a sum (not a

product) of activities. The presented analysis can be applied then as well, i.e. directly on the data. In such a case the regression coefficients β would have direct interpretation as emission coefficients (also known as emission factors). In our study no information on this kind of activities was available, and therefore a production function of known factors has been applied.

This model might be possibly extended for the case, in which the CAR prior is used for parameter coefficients β. In the context of emission inventories, this approach might be useful when considering space-varying emission factors. These models are mentioned for instance in (Gamerman and Lopes, 2006), while (Gamerman et al., 2003) provide computational details for sampling schemes in a relevant algorithm required for parameter estimation.

References

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall: London

Besag J (1974) Spatial interactions and the statistical analysis of lattice systems (with discussion). Journal of the Royal Statistical Society Series B 36:192-236

Bun R et al (2007) Spatial GHG inventory on regional level: Accounting for uncertainty. In: Proc 2nd Int Workshop Uncertainty in Greenhouse Gas Inventories.IIASA, Laxenburg, 27-32

Cressie N (1993) Statistics for spatial data. Revised edition, Wiley

Gamerman D, Moreira ARB, Rue H (2003) Space-varying regression models: specifications and simulation. Computational Statistics & Data Analysis 42:513-533

Gamerman D, Lopes HF (2006) Markov chain Monte Carlo. Stochastic simulation for Bayesian inference. 2nd edition. Chapman and Hall: London

Kaiser MS et al (2002) Analysis of particulate matter air pollution using Markov random field models of spatial dependence. Environmetrics 13:615-628

Liddle AR (2007) Information criteria for astrophysical model selection. Monthly Notices of the Royal Astronomical Society: Letters 377(1):L74-L78(1)

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325-337

Oijen M van, Thomson A (2007) Towards Bayesian uncertainty quantification for forestry models used in the U.K. GHG inventory for LULUCF. In: Proc 2nd Int Workshop Uncertainty in Greenhouse Gas Inventories.IIASA, Laxenburg, 209-216

Perez-Ramirez J (2006) Prospects of N_2O emission regulations in the European fertilizer industry. Applied Catalysis B: Environmental 70:31-35

Pope VD, Gallani M, Rowntree PR, Stratton RA (2000) The impact of new physical parametrizations in the Hadley Centre climate model - HadAM3. Climate Dynamics 16:123-146

Press J (2003) Subjective and Objective Bayesian Statistics. Second edition, Wiley

R Development Core Team (2008) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at (www.r-project.org) ISBN 3-900051-07-0

Robert ChP, Casella G (2004) Monte Carlo statistical methods. Second Edition, New York: Springer

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society Series B 64:583-639

Theloke J et al (2007) Development of a methodology for temporal and spatial resolution of greenhouse gas emission inventories for validation. In: Proc 2nd Int Workshop Uncertainty in Greenhouse Gas Inventories.IIASA, Laxenburg, 203-208

Winiwarter W, Dore Ch, Hayman G et al (2003) Methods for comparing gridded inventories of atmospheric emissions - application for Milan province, Italy and the Greater Athens Area, Greece. The Science of the Total Environment 303:231-243

Winiwarter W (2007) Personal communication