INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

TECHNIKI INFORMACYJNE TEORIA I ZASTOSOWANIA

Wybrane problemy
Tom 3 (15)

poprzednio

ANALIZA SYSTEMOWA W FINANSACH
I ZARZĄDZANIU

Pod redakcją
Andrzeja MYŚLIŃSKIEGO

Warszawa 2013

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

TECHNIKI INFORMACYJNE
TEORIA I ZASTOSOWANIA

Wybrane problemy
Tom 3 (15)

poprzednio

**ANALIZA SYSTEMOWA W FINANSACH
I ZARZĄDZANIU**

Pod redakcją
Andrzeja Myślińskiego

Warszawa 2013

Wykaz opiniodawców artykułów zamieszczonych
w niniejszym tomie:

Dr hab. inż. Maria GANZHA, prof. PAN

Prof. dr hab. inż. Zbigniew NAHORSKI

Dr hab. Marcin PAPRZYCKI, prof. PAN

Prof. dr hab. inż. Andrzej STRASZAK

Prof. dr hab. inż. Stanisław WALUKIEWICZ

ISBN 9788389475442

SPATIAL DISAGGREGATION OF AIR POLLUTION DATA WITH CONDITIONAL AUTOREGRESSIVE MODEL

Joanna Horabik

Studia Doktoranckie IBS PAN, Instytut Badañ Systemowych PAN,
e-mail: *Joanna.Horabik@ibspan.waw.pl*

Abstract.. The purpose of this study is to develop a method for allocating pollutant concentrations to finer spatial scales conditional on covariate information observable in a fine grid. Spatial dependence is modeled with the conditional autoregressive structure. The maximum likelihood approach to inference is employed, and the optimal predictors are developed to assess missing concentrations in a fine grid. The method is developed for a practical application of an output from the dispersion model CALPUFF run for Warsaw agglomeration.

Key words: air pollution, conditional autoregressive model, disaggregation, spatial correlation

1 INTRODUCTION

Atmospheric dispersion models constitute a basic tool for air quality control. Further usage of output from dispersion models include, among others, health impact assessments. For improved risk assessments, it is often required to develop air quality data in a resolution higher than the one readily available from dispersion models.

Making inference on variables at points or grid cells different from those of the data is referred to as *the change of support problem*. Several approaches have been proposed to address the issue. The geostatistical solution for realignement from point to areal data is provided by block kriging [4, 5]. In the case that data are observed at areal units and inference is sought at a new level of spatial aggregation, areal weighting offers a straightforward approach. Some improved approaches with better covariate modeling were also proposed e.g. in [12, 13].

In the following we present an approach for areal to areal data realignement; it accounts for a tendency toward spatial clustering and is focused on application to air quality. The approach resembles to some extent the

method originally proposed in [2] for disaggregation of time series based on related, higher frequency series; see also [14]. Here, a similar methodology is employed to disaggregate spatially correlated data. Regarding an assumption on residual covariance, we apply the structure suitable for area data, i.e. the conditional autoregressive (CAR) model. Although the CAR specification is typically used in epidemiology [1], it was also successfully applied for modelling air pollution over space [9, 11]. Compare also [8] for another application of the CAR structure to model spatial inventory of GHG emissions.

For inference the maximum likelihood approach is employed, and the optimal predictors are developed to assess missing concentrations in a fine grid. We demonstrate usefulness of the disaggregation method for an output from the dispersion model CALPUFF run for Warsaw agglomeration.

2 THE MOTIVATING DATA SET

The study concerns air pollution concentrations obtained from the dispersion model CALPUFF v.5 (Earthtech, Inc.) [15]. CALPUFF is a regional model for the simulation of atmospheric pollution transportation. It accounts for atmospheric processes, including their variability in space and time. The meteorological characteristic is generated by the CALMET module which, among others, accounts for the impact of terrain shape [10].

The study concerns Warsaw agglomeration within the administrative borders. The CALPUFF model, run for the real emission and meteorological data from 2005, provided atmospheric concentrations for 16 kinds of air pollutants, among others fine particulates PM_{10} and $PM_{2.5}$, nitrogen oxides NO_x, sulphur dioxide SO_2, and heavy metals. Four categories of emission sources were considered: high and low point emission sources, as well as area and line emission sources [6, 7].

The proposed disaggregation method is suitable for air pollution concentrations arising from line emission sources (transportation), and it is less adequate in the case of point emission sources (energy and industry). City roads (Fig. 1) are the main emission sources for NO_x and PM_{10}, unlike for instance SO_2; compare the scheme in Fig. 2 generated from the considered CALPUFF output. Therefore, the proposed disaggregation method was tested for concentration of NO_x and PM_{10}. The results presented in this manuscript concern NO_x concentrations; the results obtained for particulate matter PM_{10} were very similar.

The CALPUFF output for Warsaw area is generated in the 1 km grid, comprising 563 grid cells. Since the main goal of the present study is to

Fig. 1. Road communication network in Warsaw agglomeration

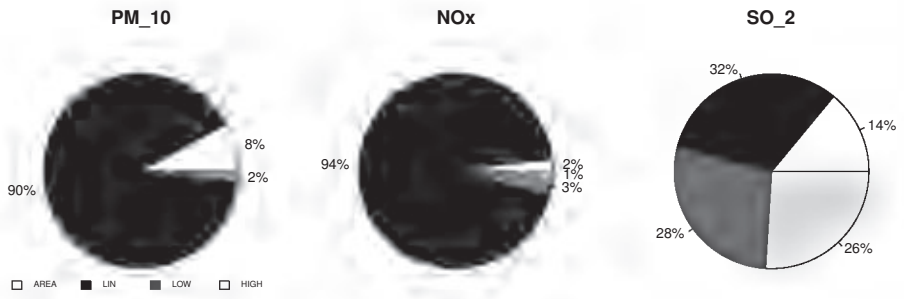

Fig. 2. Contribution of emission categories (AREA - area; LIN - line; LOW - low point sources; HIGH - high point sources) in the concentration of PM_{10}, NO_x and SO_2

verify the proposed disaggregation procedure, therefore we treat the model output as a reference data set. Based on these values we calculate NO_x concentration in the 2 km grid, assuming that the contribution of each 1 km cell is proportional to area. Next, we apply the proposed model for disaggregation of 2 km grid data back into 1 km grid; the resulting values are subsequently compared with the original concentrations. The maps of NO_x concentrations in the 1 km grid (the CALPUFF output) and in 2 km grid

(values to be disaggregated) are presented in Fig. 3; throughout the paper the pollutant concentrations are reported in $\mu g/m^3$.

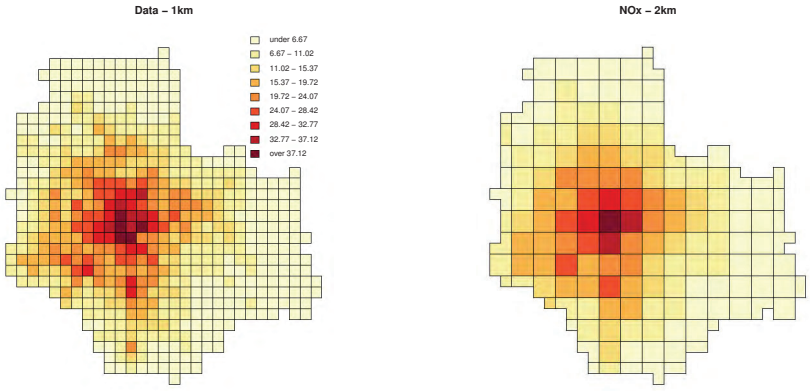

Fig. 3. NO_x concentration: original CALPUFF output in 1 km grid (left), and the values to be disaggregated in 2 km grid (right)

The obtained results were compared with the geostatistical approach. The conditional autoregressive model is an alternative for the geostatistical approach, which refers to point data, compare e.g. [3]. Within the geostatistical model, the spatial dependence is characterised by means of a distance between data points, while the CAR structure is based on the notion of neighbourhood between areas (grid cells).

In the study, the sensitivity of results with respect to a neighbourhood structure was also assessed. The performance of the disagrregation procedure was compared for the models of 4 neighbours (the so-called Rook Method; denoted CAR4) and for the ones of 8 neighbours (the so-called Queen Method; denoted CAR8), see Fig. 4.

3 THE DISAGGREGATION MODEL

We begin with the model specification in a fine 1 km grid. Let Y_i denote a random variable associated with a missing value of pollutant, here NO_x level, y_i defined at each cell i, $i = 1, ..., n$ of a fine grid. Assume that random variables Y_i follow a Gaussian distribution with the mean μ_i and variance σ_Y^2,

$$Y_i|\mu_i \sim \mathcal{N}\left(\mu_i, \sigma_Y^2\right). \tag{1}$$

Fig. 4. Neighbourhood scheme for models CAR4 and CAR8

Given the values μ_i and σ_Y^2, the random variables Y_i are independent.

The values $\boldsymbol{\mu} = \{\mu_i\}_{i=1}^n$ represent the true process underlying NO_x concentration, and the (missing) observations are related to this process through a measurement error of variance σ_Y^2. The model for the underlying NO_x process is formulated as a sum of regression component with available covariates, and a spatially varying random effect.

The approach to modelling μ_i expresses an assumption that available covariates explain part of the spatial pattern, and the remaining part is captured through a spatial dependence, introduced as the conditionally autoregressive structure. The CAR scheme follows an assumption of similar random effects in adjacent cells, and it is given through the specification of full conditional distribution functions of μ_i for $i = 1, \ldots, n$ (see e.g. [1, 3, 4])

$$\mu_i|\boldsymbol{\mu}_{-i} \sim \mathcal{N}\left(\boldsymbol{x}_i^T\boldsymbol{\beta} + \rho\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{w_{ij}}{w_{i+}}\left(\mu_j - \boldsymbol{x}_j^T\boldsymbol{\beta}\right), \frac{\tau^2}{w_{i+}}\right), \tag{2}$$

where $\boldsymbol{\mu}_{-i}$ denotes all elements in $\boldsymbol{\mu}$ but μ_i; w_{ij} are the adjacency weights ($w_{ij} = 1$ if j is a neighbour of i and 0 otherwise, also $w_{ii} = 0$); w_{i+} is the number of neighbours of area i; $\boldsymbol{x}_i^T\boldsymbol{\beta}$ is a regression component with explanatory covariates for area i and a respective vector of regression coefficients; and τ^2 is a variance parameter. Thus, the mean of the conditional distribution $\mu_i|\boldsymbol{\mu}_{-i}$ consists of the regression part and the second summand, being proportional to the average values of remainders $\mu_j - \boldsymbol{x}_j^T\boldsymbol{\beta}$ for neighbouring sites (i.e. when $w_{ij} = 1$). The proportion is calibrated with the parameter ρ, reflecting strength of spatial correlation. Furthermore, the variance of the conditional distribution (2) is inversely proportional to the number of neighbours w_{i+}.

The joint distribution of the process $\boldsymbol{\mu}$ is the following (for derivation see e.g. [3, 4])

$$\boldsymbol{\mu} \sim \mathcal{N}_n \left(\boldsymbol{X}\boldsymbol{\beta}, \tau^2 \left(\boldsymbol{D} - \rho \boldsymbol{W} \right)^{-1} \right), \tag{3}$$

where $\boldsymbol{X}$ is a design matrix with vectors $\boldsymbol{x}_i$; $\boldsymbol{D}$ is an $n \times n$ diagonal matrix with w_{i+} on the diagonal; and $\boldsymbol{W}$ is an $n \times n$ matrix with adjacency weights w_{ij}. Equivalently, we can write (3) as

$$\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \;\; \boldsymbol{\epsilon} \sim \mathcal{N}_n \left(\boldsymbol{0}, \boldsymbol{N} \right) \tag{4}$$

with $\boldsymbol{N} = \tau^2 \left(\boldsymbol{D} - \rho \boldsymbol{W} \right)^{-1}$.

The model for the CALPUFF output data, observed in a 2 km grid, is obtained by multiplication of the mean process $\boldsymbol{\mu}$ with an $N \times n$ *aggregation matrix* C, where N is a number of observations in a 2km grid

$$\boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\epsilon}, \;\; \boldsymbol{C}\boldsymbol{\epsilon} \sim \mathcal{N}_N \left(\boldsymbol{0}, \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T \right). \tag{5}$$

The matrix C consists of 0's and 1's, indicating which cells have to be aligned together. The random variable $\boldsymbol{\lambda} = \boldsymbol{C}\boldsymbol{\mu}$ is treated as the mean process for variables $\boldsymbol{Z} = \{Z_i\}_{i=1}^N$ associated with observations $\boldsymbol{z} = \{z_i\}_{i=1}^N$ of the aggregated model

$$\boldsymbol{Z}|\boldsymbol{\lambda} \sim \mathcal{N}_N \left(\boldsymbol{\lambda}, \sigma_Z^2 \boldsymbol{I}_N \right), \tag{6}$$

where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. Also at this level, the underlying process $\boldsymbol{\lambda}$ is related to $\boldsymbol{Z}$ through a measurement error with variance σ_Z^2.

The parameters $\boldsymbol{\beta}, \sigma_Z^2, \tau^2$ and ρ are estimated with the maximum likelihood method based on the joint unconditional distribution

$$\boldsymbol{Z} \sim \mathcal{N}_N \left(\boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T \right),$$

where $\boldsymbol{M} = \sigma_Z^2 \boldsymbol{I}_N$. The analytical derivation is limited to the regression coefficients $\boldsymbol{\beta}$, and further maximisation of the profile log likelihood is performed numerically. The standard errors of estimators are calculated with the expected Fisher information matrix.

Regarding the missing values in a fine 1km grid, the underlying NO_x process is of our primary interest. The predictors optimal in terms of the minimum mean squared error are given by $E\left(\boldsymbol{\mu}|\boldsymbol{z}\right)$. The joint distribution of $(\boldsymbol{\mu}, \boldsymbol{Z})$ is

$$\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{Z} \end{bmatrix} \sim \mathcal{N}_{n+N} \left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{N} & \boldsymbol{N}\boldsymbol{C}^T \\ \boldsymbol{C}\boldsymbol{N} & \boldsymbol{M} + \boldsymbol{C}\boldsymbol{N}\boldsymbol{C}^T \end{bmatrix} \right). \tag{7}$$

The distribution (7) allows for full inference, yielding both the predictor and its error

$$\widehat{E(\boldsymbol{\mu}|\boldsymbol{z})} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{N}}\boldsymbol{C}^T\left(\widehat{\boldsymbol{M}} + \boldsymbol{C}\widehat{\boldsymbol{N}}\boldsymbol{C}^T\right)^{-1}\left[\boldsymbol{z} - \boldsymbol{C}\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right]$$
$$\widehat{Var(\boldsymbol{\mu}|\boldsymbol{z})} = \widehat{\boldsymbol{N}} - \widehat{\boldsymbol{N}}\boldsymbol{C}^T\left(\widehat{\boldsymbol{M}} + \boldsymbol{C}\widehat{\boldsymbol{N}}\boldsymbol{C}^T\right)^{-1}\boldsymbol{C}\widehat{\boldsymbol{N}}.$$

Note that in the predictor $\widehat{E(\boldsymbol{\mu}|\boldsymbol{z})}$, a naive regression forecast is corrected with a residual on the aggregated grid distributed over respective grid cells.

4 RESULTS

Table 1 presents estimation results (parameter estimates and their standard errors) for two conditional autoregressive models (CAR4 and CAR8), for the geostatistical one (GEOST) and for linear regression (LM). The listed approaches were analysed either with the trend based on three categories of roads, or without trend (with a constant value). Within the last group, we also considered the naive prediction (NAIVE), where equal values are assumed for each 1 km$\times$1 km grid cell within a respective 2 km$\times$2 km area.

For all the models with trend, the ratio of regression coefficients and their respective standard errors (i.e. the t-test statistics) roughly indicate that all the considered types of roads are statistically significant. This is not true anymore for β_0. As to be expected, insignificant coefficient β_0 is reported for the models CAR4, CAR8 and GEOST, thus those with a spatial component.

For the models CAR4 and CAR8, the differences among respective parameter values are negligible. Note, however, that when a higher number of neighbours is accounted for (model CAR8), the parameter ρ is lower (0.9994) and the variance related with a spatial component (τ^2) is higher.

When it comes to the geostatistical model, the exponential model of semivariogram was applied, see for instance [4]. The parameter σ_Z^2 denotes the variogram sill, the parameter ϕ stands for the semivariogram range, and the nugget effect is absent in this case study.

The (negative) loglikelihood values ($-L$) as well as the Akaike criterion (AIC) are reported in Table 2. The geostatistical approach provides the best fit to the data in 2 km grid (the lowest values of $-L$ and AIC), both in the group of models with and without trend.

The values predicted in the 1 km grid by means of the models with trend are shown in Fig. 5, while Fig. 8 shows those resulting from the

Table 1. Maximum likelihood estimates for 2km grid

	CAR4		CAR8		GEOST		LM	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
Models with trend								
β_0	6.754	4.383	6.652	4.500	3.361	3.16	2.217	0.557
β_1	0.0019	2.59e-04	0.002	2.64e-04	0.0017	2.37e-04	0.0041	4.87e-04
β_2	0.0014	2.93e-04	0.0016	3.01e-04	0.0012	2.70e-04	0.0047	5.31e-04
β_3	0.0005	5.45e-05	0.0005	5.62e-05	0.0005	5.15e-05	0.0009	8.72e-05
σ_Z^2	3.65e-08	0.296	0.296e-08	0.338	25.87	9.82	15.375	1.735
τ^2	18.56	2.095	51.88	5.86	-	-	-	-
ρ	0.9995	3.29e-09	0.9994	4.88e-09	-	-	-	-
ϕ	-	-	-	-	17 105	6 477	-	-
Models without trend								
β_0	11.717	6.204	11.814	6.247	1.826	5.47	-	-
σ_Z^2	7.33e-07	0.590	1.59e-07	0.671	63.33	15.93	-	-
τ^2	36.89	4.16	103.03	11.62	-	-	-	-
ρ	0.9995	2.28e-09	0.9994	3.58e-09	-	-	-	-
ϕ	-	-	-	-	23 215	5 404	-	-

Table 2. Model comparison

	$-L$	AIC
Models with trend		
CAR4	349.7	713.4
CAR8	353.9	721.9
GEOST	323.9	659.8
LM	437.3	884.6
Models without trend		
CAR4	403.7	815.3
CAR8	407.8	823.6
GEOST	371.1	748.2

models without trend. The differences between the obtained values and the original data (compare Fig. 3) can be noticed easily. In the case of the models with trend, both the approaches CAR4 and CAR8 provide quite good predictions. Similarly it is for the geostatistical model, although high NO_x concentrations tend to be underestimated in this case. This observation is confirmed in the scatterplot (Fig. 7). In the model GEOST without trend this feature is even more pronounced (see Fig. 9).

The linear regression (LM) model is completely not adequate; this is clear from the map of predicted concentrations (Fig. 5) and from the scatterplot (Fig. 7). Also, in Fig. 6 the map of residuals for two chosen models with trend (CAR4 and LM) illustrates importance of the spatial aspect in this study. On the other hand, the models without trend do not perform

well either, proving that also information on the road network is essential for accurate predictions.

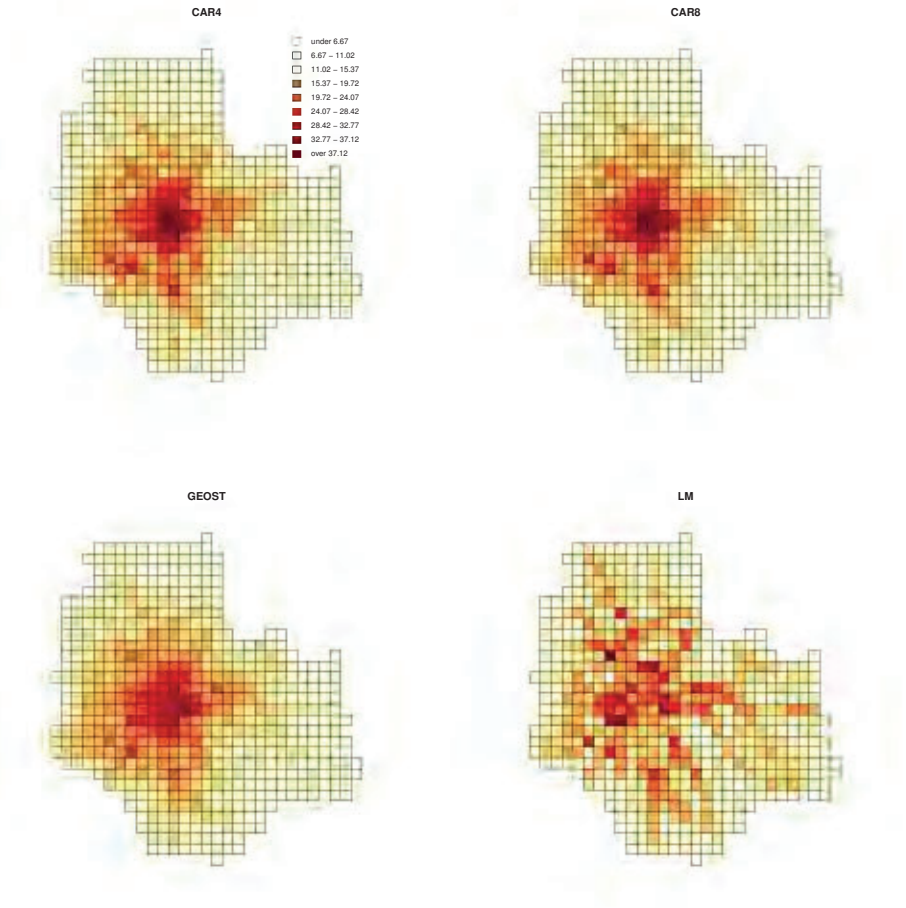

Fig. 5. NO_x concentrations predicted in 1km grid - the models with trend

Further analysis of residuals ($d_i = y_i - y_i^*$, where y_i^* - predicted values) for all the models is presented in Table 3. It comprises the mean squared error mse, the minimum and maximum values of d_i as well as the sample correlation coefficient r between the predicted y_i^* and observed y_i values. Within the group of models with trend, the conditionally autoregressive model CAR4 and geostatistical one provide the best predictions: the mean squared error mse=2.9, the correlation coefficient r=0.977, also the same

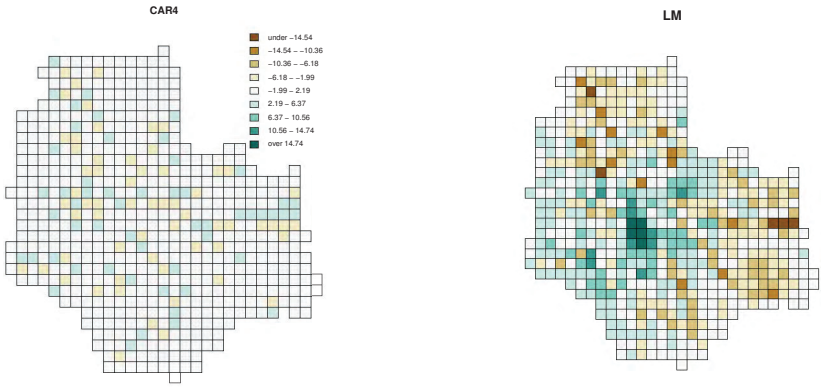

Fig. 6. Residuals from CAR4 model with trend (left), and from linear regression LM (right)

values of min(d_i)=-5.95 are reported. Note, that despite a slightly worse fit of CAR4 model to the aggregated data (compare Table 2), its predictive performance is as good as that of GEOST model.

Within the group of models without trend, the models CAR4 and CAR8 provide the best predictions. Therefore, in the lack of covariate information the proposed procedure based on the conditional autoregressive scheme outperforms the GEOST model. Interestingly, due to high spatial correlation of the pollutant concentration, the NAIVE approach delivers far better predictions than the linear regression (LM); compare Table 3 and respective graphs in Fig. 7 and Fig. 9.

Table 3. Analysis of residuals ($d_i = y_i - y_i^*$)

	mse	min(d_i)	max(d_i)	r
Models with trend				
CAR4	2.97	-5.95	5.78	0.977
CAR8	3.09	-6.00	5.77	0.976
GEOST	2.93	-5.95	6.69	0.977
LM	29.75	-18.72	18.92	0.769
Models without trend				
CAR4	3.84	-6.52	8.30	0.970
CAR8	3.85	-6.49	8.39	0.970
GEOST	4.57	-7.01	9.57	0.964
NAIVE	5.26	-7.19	8.29	0.959

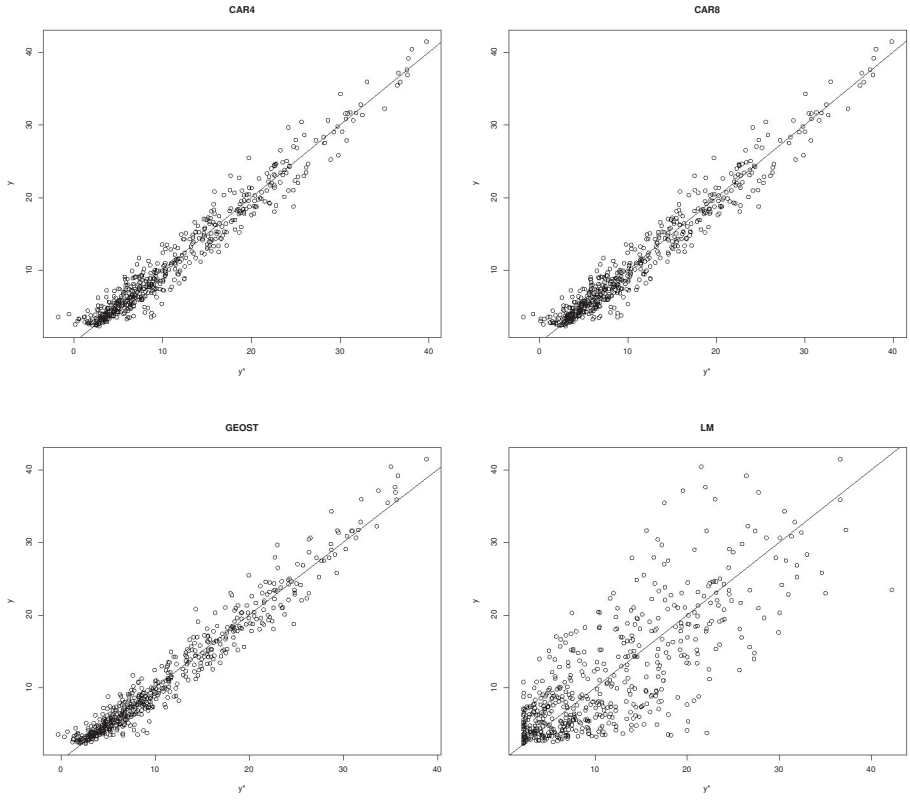

Fig. 7. Scatterplots of predicted (y_i^*) versus observed (y_i) values - the models with trend

5 DISCUSSION AND CONCLUDING REMARKS

The major objective of this study was to present a method for allocating pollutant concentrations to finer spatial scales conditional on covariate information observable in a fine grid. Spatial dependence was modeled with the conditional autoregressive structure, which was introduced as a random effect. The maximum likelihood approach to inference was employed providing parameter estimates, and their standard errors were also obtained. Furthermore, we developed the optimal predictors to assess missing concentrations in a fine grid, along with the estimates of prediction error.

The proposed scheme was verified on the output from pollution dispersion model CALPUFF, namely NO_x concentrations for Warsaw agglomeration. Detailed information on the city road network served as covariates,

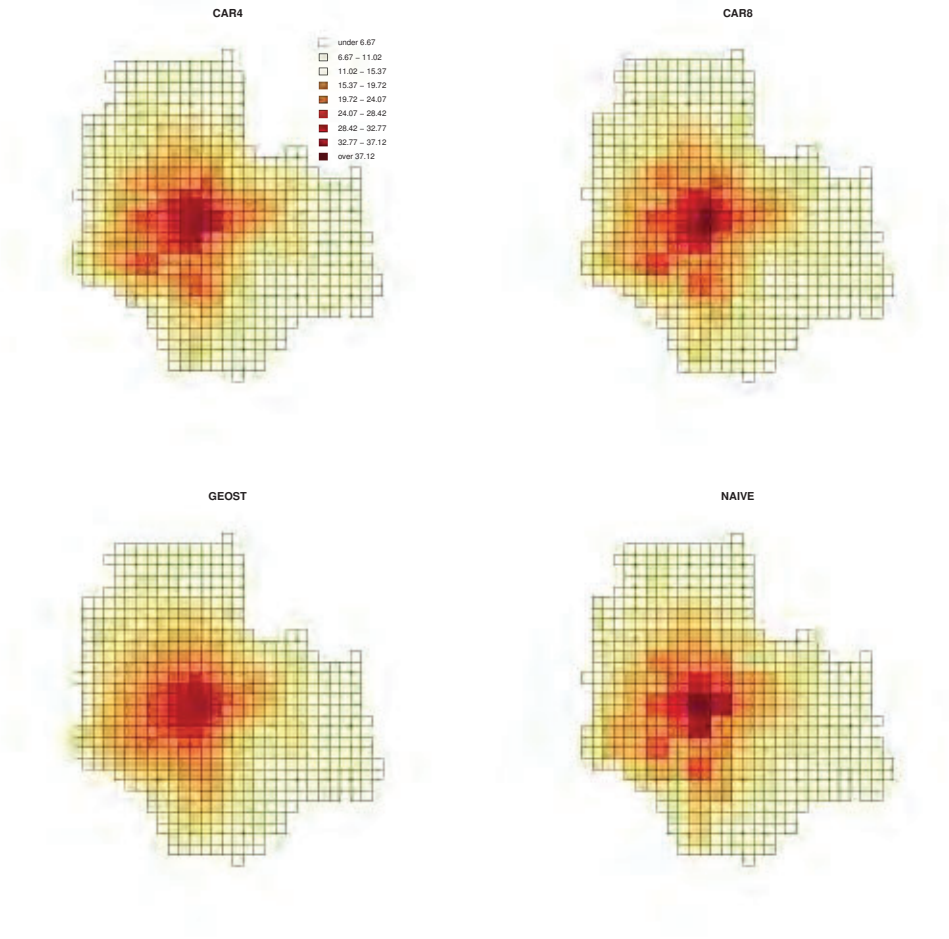

Fig. 8. NO_x concentrations predicted in the fine grid - the models without trend

since transportation is the main source of NO_x pollution. A fourfold disaggregation from 2 km regular grid into 1 km grid was tested.

With available information on the city road network, the model based on the conditionally autoregressive structure provided equally good predictions as the one based on the geostatistics. In this case the sample correlation coefficient was equal r=0.977. It is also worth noting that for the high range values, the model based on the CAR scheme provided more accurate predictions, while the geostatistical one tends to underestime these values. Correct predictions of high range NO_x and PM_{10} concentrations are especially important, since the thresholds of these pollutants (40 μg/m^3) are

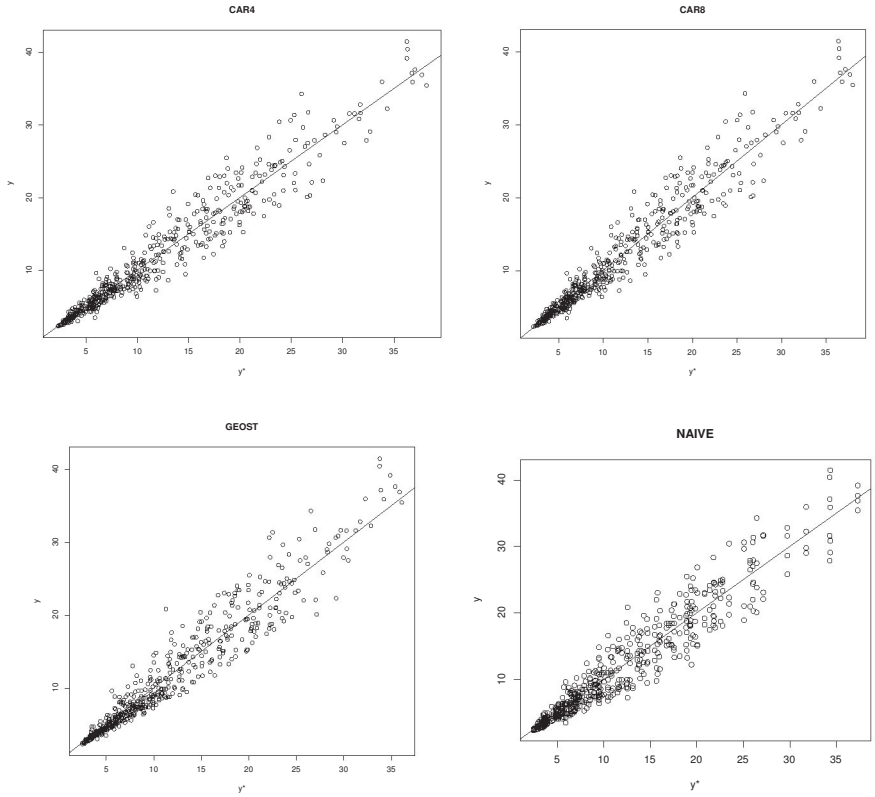

Fig. 9. Scatterplots of predicted (y_i^*) versus observed (y_i) values - the models without trend

exceeded in some parts of Warsaw agglomeration. When no information on roads was available, better results were obtained for CAR models; especially the mean squared error was lower.

The impact of the neighbourhood structure was negligible, although both in the models with and without trend slightly better results were obtained for the model CAR4 than for CAR8. Apparently, high spatial correlation is already exploited enough with four closest neighbours; additional neighbours under CAR8 model increase variability and do not add much to explanatory power. Also, the proposed model provided much better predictions than the naive approach, which assumed equal pollutant concentrations in the whole 2 km cells.

ACKNOWLEDGEMENT. This contribution is supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing; project financed from The European Union within the Innova-

tive Economy Operational Programme 2007-2013 and European Regional Development Fund.

References

1. Banerjee S., Carlin B.P., Gelfand A.E. (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC.
2. Chow G. C., Lin A. (1971) Best linear unbiased interpolation, distribution, and extrapolation of time series by related series, *The Review of Economics and Statistics*, **53**, 372-375.
3. Cressie N.A.C. (1993) *Statistics for Spatial Data*, Wiley, New York.
4. Gelfand A.E., Diggle P.J., Fuentes M., Guttorp P., Eds. (2010) *Handbook of Spatial Statistics*, Chapman & Hall/CRC.
5. Gotway C.A., Young L.J. (2002) Combining incompatible spatial data, *Journal of the American Statistical Association*, **97**, 632-648.
6. Holnicki P., Nahorski Z. (2010) Wpşyw niepewnoIJci danych emisyjnych na dokşadnoIJæ prognoz zanieczyszczeñ atmosferycznych, *Studia i Materiaşy Polskiego Towarzystwa Zarźdzania Wiedźź*, Nr 33, Bydgoszcz, 146-157.
7. Holnicki P., Nahorski Z. (2011) *Analiza niepewnoIJci w prognozowaniu rozprzestrzeniania siê zanieczyszczeñ atmosferycznych w aglomeracji warszawskiej*, Raport z prac wykonanych w ramach projektu NN 519316735, Instytut Badañ Systemowych Polskiej Akademii Nauk, Warszawa.
8. Horabik J., Nahorski Z. (2010) A statistical model for spatial inventory data: a case study of N2O emissions in municipalities of southern Norway, *Climatic Change*, **103**, Issue 1-2, 263-276.
9. Kaiser M. S., Daniels M. J., Furakawa K., Dixon P. (2002) Analysis of particulate matter air pollution using Markov random field models of spatial dependence, *Environmetrics*, **13**, 615-628.
10. Markiewicz M.T. (2004) *Podstawy rozprzestrzeniania siê zanieczyszczeñ w powietrzu atmosferycznym*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
11. McMillan N.J., Holland D.M., Morara M., Feng J. (2010) Combining numerical model output and particulate data using Bayesian space-time modeling, *Environmetrics*, **21**, 48-65.
12. Mugglin A.S., Carlin B.P. (1998) Hierarchical modeling in geographical information systems: Population interpolation over incompatible zones, *Journal of Agricultural, Biological and Environmental Statistics*, **3**, 111-130.
13. Mugglin A.S., Carlin B.P., Gelfand A.E. (2000) Fully model-based approaches for spatially misaligned data, *Journal of the American Statistical Association*, **95**, 877-887.
14. Polasek W., Llano C., Sellner R. (2010) Bayesian methods for completing data in spatial models, *Review of Economic Analysis*, **2**, 194-214.
15. Scire J.S., Strimaitis D.G., Yamartino R.J. (2000) *A User's Guide for the CALPUFF Dispersion Model*, Earth Technology Inc.

PRZESTRZENNA DEZAGREGACJA DANYCH O EMISJACH ZANIECZYSZCZEŃ ATMOSFERYCZNYCH - ZASTOSOWANIE MODELU WARUNKOWO AUTOREGRESYJNEGO

Streszczenie. W pracy zaproponowano model przestrzennej dezagregacji danych do drobnej siatki w oparciu o dane zagregowane z wykorzystaniem

dodatkowych informacji dostępnych w drobnej siatce. Do modelowania przestrzennej zależności obserwacji wykorzystano model warunkowo autoregresyjny ze względu na fakt, iż wartości w węźle siatki są utożsamiane ze średnim stężeniem emisji w całym oczku siatki; zależność przestrzenną uwzględniono jako efekt losowy. Do oceny wartości zmiennej losowej związanej z brakującą obserwacją w drobnej siatce wyznaczono predyktory oraz oszacowania ich błędów. Zaproponowany model zweryfikowano na danych pochodzących z modelu CALPUFF o stężeniu zanieczyszczeń powietrza (NO_x, PM_{10}) dla obszaru aglomeracji warszawskiej w 2005 roku.

ISBN 83-894-7550-2