51/2012

Raport Badawczy

Research Report

RB/41/2012

Spatial disaggregation of activity data for GHG inventory in agricultural sector of Poland

J. Horabik

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 Zakładu zgłaszający pracę:
Prof. zw. dr hab. inż. Zbigniew Nahorski

Warszawa 2012

SYSTEMS RESEARCH INSTITUTE
POLISH ACADEMY OF SCIENCES

Joanna Horabik

Spatial disaggregation of activity data for GHG inventory in agricultural sector of Poland

Warszawa 2012

Abstract

This report presents a novel approach for allocation of spatially correlated data, such as emission inventories, to finer spatial scales, conditional on covariate information observable in a fine grid. Spatial dependence is modelled with the conditional autoregressive structure introduced into a linear model as a random effect. The maximum likelihood approach to inference is employed, and the optimal predictors are developed to assess missing values in a fine grid. The usefulness of the proposed technique is shown for agricultural sector of GHG inventory in Poland. An example of allocation of livestock data (a number of horses) from district to municipality level is analysed. The results indicate that the proposed method outperforms a naive and commonly used approach of proportional distribution.

Keywords: GHG inventory, agricultural sector, spatial correlation, disaggregation, conditional autoregressive model

Chapter 1

Introduction

Greenhouse gas (GHG) emission inventories serve as a basic tool for verification of international treaties aimed at constraing global warming. Despite all their drawbacks and limitations [14], national GHG inventories provide invaluable information on anthropogenic emission sources, and, indirectly, on effectiveness of undertaken emission abatement measures. Constant efforts of IPCC community seek to improve the inventory procedure and to limit underlying uncertainties and imprecision [13].

Although the greenhouse gases directly are not harmful for human health, their spatial distribution is of great importance. For instance, a network of ecosystem long-term observation sites is launched across Europe to understand behavior of the global carbon cycle and greenhouse gas emissions. The activities are conducted within the Integrated Carbon Observation System infrastructure. Another approach is to develop a spatially resolved GHG inventory. All of these efforts open new opportunities for improvement of emission reduction activities, including among others attribution of sources and sinks.

The present study was conducted as a part of the 7FP Marie Curie Actions project *Geoinformation technologies, spatio-temporal approaches, and full carbon account for improving accuracy of GHG inventories*. One of the main aims of the project is to develop a spatial inventory of GHG for Poland. The task comprises estimation of GHG related activity data, which need to be spatially resolved in this case, and their corresponding emission factors. In terms of considered sectors, subsectors and separate emission source groups, the IPCC guidelines [11] provide relevant methodology, and it is followed throughout the project. The main GHG emission sectors include energy (fossil fuel burning from stationary and mobile sources), industry and agriculture.

Development of spatial GHG inventory crucially depends on availability of low resolution activity data. In Poland, relevant information needs to be acquired from national/regional totals. A procedure of allocation into smaller spatial units (like districts, municipalities and finally 2x2km grid) differs among various emission sectors. Basically, all the emission sources are categorised as line, area or large point emission sources; further steps differ significantly for each group. For large point sources, such as power/heat stations or refinery plants, corresponding emissions are associated directly with a particular object located in space. Line sources, like roads, railways or pipelines, are usually analyzed by cutting line objects into sections using respective grids. Area sources comprise e.g. agricultural fields, urban areas as well as highly dense urban transportation network. In this case, a procedure of spatial allocation depends on methods and technologies of fossil fuel combustion in a considered sector [2]. A common approach though is a spatial allocation made in a proportion to some related indicators, i.e. proxy data, which are available in a finer grid. This solution to a large extent relies on subjective assumptions, and usually there is no mean for verification of the results obtained.

Within the project Work Package 3, the statistical scaling methods are developed in order to support the procedure of compiling high resolution activity data. In this report we propose the method for allocating GHG activity data to finer spatial scales conditional on covariate information, such as land use, observable in a fine grid. The proposition is suitable for spatially correlated, area emission sources.

The approach resembles to some extent the method of Chow and Lin (1971) [3], originally proposed for disaggregation of time series based on related, higher frequency series. 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 [12], [15]. Compare also [9] for another application of the CAR structure to model spatial inventory of GHG emissions. The maximum likelihood approach to inference is employed, and the optimal predictors are developed to assess missing concentrations in a fine grid. We demonstrate usefulness of the disaggregation method for spatially correlated area sources, in particular for agricultural sector.

A part of the methododology described in section 3.1 was already presented in [10]. This contribution extends the basic model for the case of various regression models in each region (here voivodeship); see section 3.2. Performance of the method for livestock data in agricultural sector of GHG inventory is presented in chapter 4.

Chapter 4

Results

First, Table 4.1 presents estimation results (parameters with their standard errors) for models with and without a spatial component, denoted CAR and LM respectively. Note that β_2 - land use class Arable land turned to be statistically insignificant in this setting. Introducing spatial CAR structure increases standard error of estimated parameters, as compared with LM model. However, for assessment of goodness of fit for these models Table 4.2 should be referred to.

Table 4.1: Maximum likelihood estimates

	CAR		LM	
	Est.	Std.Err.	Est.	Std.Err.
β_0	8.525	0.1605	-6.981	0.0389
β_1	3.517	0.0148	1.932	0.0042
β_2	-	-	-	-
β_3	0.916	0.0034	1.786	0.0010
β_4	3.912	0.0055	5.032	0.0013
σ_Z^2	0.961	0.4052	1.506	0.1202
τ^2	1.683	0.1569	-	-
ρ	0.9889	2.62e-06	-	-

Table 4.2 contains the analysis of residuals ($d_i = y_i - y_i^*$, where y_i^* - predicted values) for considered models. We report the mean squared error mse, the minimum and maximum values of d_i as well as the sample correlation coefficient r between the predicted and observed values. From here it is obvious that the spatial CAR structure considerably improve the results obtained with the model of independent errors LM. For comparison, we also include the results obtained with an allocation done proportionally to population in municipalities; this approach is called NAIVE. It is a straightforward, commonly used approach in this area of application. Here we note that the NAIVE approach provides reasonable results, but CAR model outperforms it in terms of all the reported criteria. The decrease of the mean squared error is from 3374.4 for NAIVE to 3069.4 for CAR, which gives 9% improvement.

From the maps of predicted values for the models CAR and NAIVE (Figure 4.1) it is difficult to spot a meaningful difference. The map of residuals (Figure 4.2) and scatterplot

(Figure 4.3) are slightly more informative.

Table 4.2: Analysis of residuals ($d_i = y_i - y_i^*$)

	mse	$\min(d_i)$	$\max(d_i)$	r
CAR	3069.4	-275	469	0.784
LM	5641.2	-357	522	0.555
CAR*	3437.0	-258	512	0.763
LM*	4876.1	-374	546	0.651
CAR**	3124.9	-256	446	0.783
LM**	4427.6	-352	472	0.674
NAIVE	3374.4	-475	403	0.766

Next, we considered the models with various regression coefficients in voivodeships but having the same same set of covariates (β_0,β_1, β_3 and β_4); the models are denoted CAR* and LM*, respectively. Note that the model CAR* gives much worse results than the models CAR and NAIVE.

Further, considered were the models with varying across regions both the coefficients and sets of covariates. Only statistically significant covariates were chosen. Table 4.3 includes regression coefficients along with their standard errors for all the considered regions (voivodeships), indexed with l. A reference list with voivodship names is included in the Appendix.

Table 4.3: Maximum likelihood estimates of the models CAR** and LM**

	CAR**		LM**		CAR**		LM**	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
	l=1				l=2			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	3.514	0.0528	1.289	0.0098	5.227	0.0592	3.431	0.0099
β_2^l	-	-	-	-	-	-	-	-
β_3^l	1.593	0.0221	2.063	0.0060	0.588	0.0194	1.032	0.0044
β_4^l	1.344	0.0322	3.049	0.0052	4.759	0.0288	2.909	0.0048
$(\sigma_Z^l)^2$	1.281	1.1759	0.559	0.1552	1.0905	1.6542	0.368	0.1194
	l=3				l=4			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	23.849	0.0966	24.729	0.0331	-3.349	0.0967	-2.611	0.0301
β_2^l	-1.546	0.0085	-1.679	0.0033	-	-	-	-
β_3^l	4.632	0.0196	4.308	0.0043	3.056	0.0164	2.447	0.0043
β_4^l	1.622	0.0187	2.119	0.0051	6.271	0.0512	5.129	0.0150
$(\sigma_Z^l)^2$	0.974	2.2569	2.616	0.8273	0.852	1.7905	0.614	0.2509
	l=5				l=6			

Table 4.3: (continued)

	CAR**		LM**		CAR**		LM**	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
β_0^l	-	-	-	-	-	-	-	-
β_1^l	6.392	0.0678	6.409	0.0272	0.729	0.0407	-2.221	0.0122
β_2^l	-	-	-	-	-	-	-	-
β_3^l	-	-	-	-	1.662	0.0205	4.276	0.0066
β_4^l	1.726	0.0253	2.122	0.0117	4.080	0.0199	5.117	0.0062
$(\sigma_Z^l)^2$	0.938	1.6488	2.0944	0.6463	1.382	2.7181	2.723	0.8835
	l=7				l=8			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	2.332	0.0348	4.452	0.0250	3.739	0.0648	3.491	0.0145
β_2^l	-	-	-	-	-	-	-	-
β_3^l	-	-	-	-	0.731	0.0438	0.489	0.0122
β_4^l	7.698	0.0148	8.459	0.0111	-	-	-	-
$(\sigma_Z^l)^2$	1.127	1.4045	7.5264	1.749	0.955	2.134	0.640	0.2731
	l=9				l=10			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	-	-	-	-	-	-	-	-
β_2^l	0.652	0.0078	0.686	0.0021	0.956	0.0038	0.897	0.0013
β_3^l	2.543	0.0166	1.865	0.0056	-	-	-	-
β_4^l	3.660	0.0157	3.135	0.0039	2.857	0.0101	4.322	0.0035
$(\sigma_Z^l)^2$	1.227	1.7052	0.998	0.3080	0.809	2.1353	2.145	0.8106
	l=11				l=12			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	11.063	0.0655	14.421	0.0200	2.562	0.0543	1.170	0.0097
β_2^l	-0.456	0.0045	-0.625	0.0013	0.1315	0.0097	0.523	0.0013
β_3^l	-	-	-	-	-	-	-	-
β_4^l	5.397	0.0163	4.034	0.0053	2.595	0.0390	2.142	0.0069
$(\sigma_Z^l)^2$	1.139	1.8027	1.301	0.4602	1.016	2.6822	0.636	0.2182
	l=13				l=14			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	-	-	-	-	16.235	0.0585	14.090	0.0318
β_2^l	-0.114	0.0056	-0.073	0.0021	-	-	-	-
β_3^l	-	-	-	-	-	-	-	-
β_4^l	7.445	0.0229	7.368	0.0070	1.569	0.0147	3.273	0.0107
$(\sigma_Z^l)^2$	0.515	1.7805	1.735	0.6805	0.858	1.1953	3.189	1.0349
	l=15				l=16			
β_0^l	-	-	-	-	-	-	-	-
β_1^l	2.367	0.0312	2.001	0.0100	13.159	0.0630	10.993	0.0189
β_2^l	0.615	0.0031	0.458	0.0012	-	-	-	-
β_3^l	1.652	0.0095	1.793	0.0038	-	-	-	-
β_4^l	-	-	-	-	0.379	0.0237	-0.160	0.0089
$(\sigma_Z^l)^2$	0.627	0.993	1.303	0.3311	0.634	1.4092	1.018	0.339

Table 4.3: (continued)

	CAR**		LM**		CAR**		LM**	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
τ^2	**1.647**	0.1536	-	-				
ρ	0.9913	**1.59e-06**	-	-				

The reported values of estimated parameters for CAR** and LM** show considerable differences across the voivodeships, not only in terms of estimated values of regression coefficients, but also in terms of their significance. Moreover, from Table 4.2 we note that this setting (CAR**) provides comparable results to CAR.

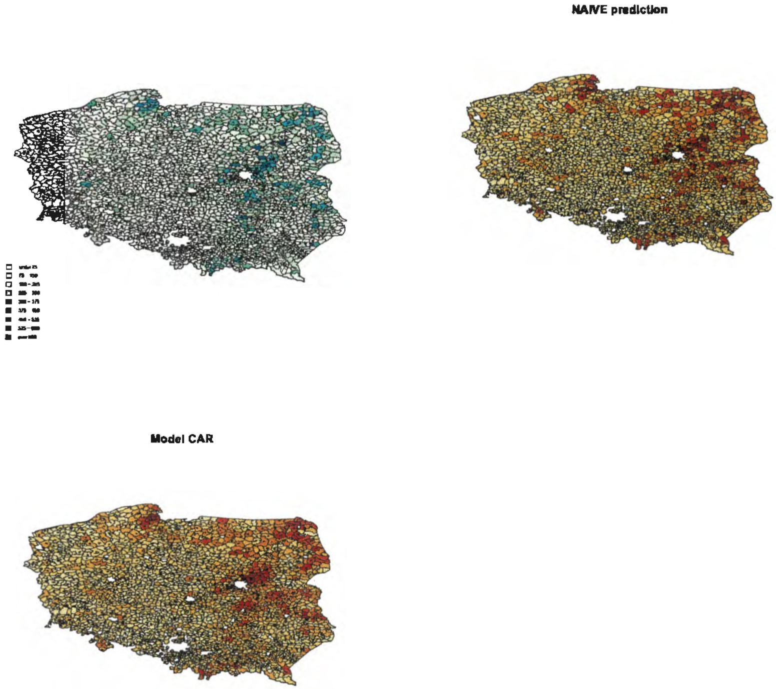

Figure 4.1: Original data in municipalities and predicted values for the models NAIVE and CAR

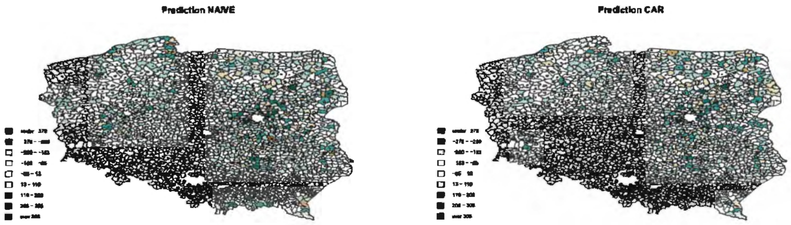

Figure 4.2: Residuals from predicted values for the models NAIVE and CAR

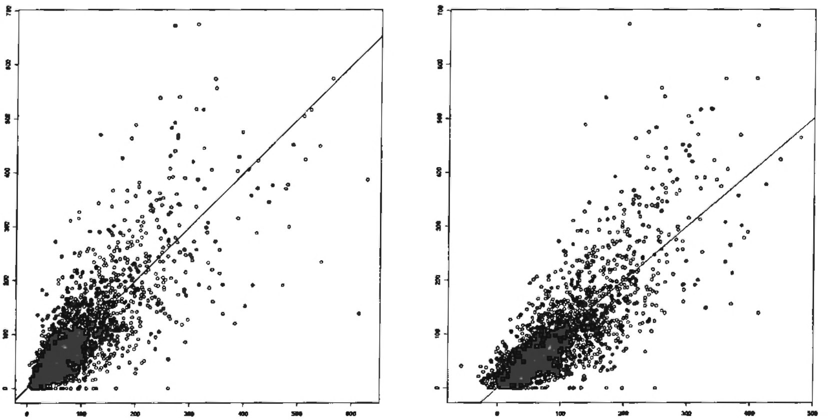

Figure 4.3: Scatterplot of predictions (y_i^*) against observations (y_i) for the models NAIVE (left) and CAR (right)

Chapter 5

Concluding remarks and discussion

The study presents the first attempt to apply the spatial scaling model for the GHG inventory in Poland. The task was to allocate spatially correlated data to finer spatial scales, conditional on covariate information observable in a fine grid. Spatial dependence is set and it is assummed not to change with the change of grid. It is modelled with the conditional autoregressive structure introduced into a linear model as a random effect. The maximum likelihood approach to inference is employed, and the optimal predictors are developed to assess missing values in a fine grid. The usefulness of the proposed technique is shown on an example of allocation of livestock data (a number of horses) from district to municipality level.

The results of the disaggregation with the proposed procedure were compared with the allocation proportional to population of municipalities. An improvement over the naive, proportional approach of 9% in terms of the mean squared error was reported. In addition, we extended the model to allow for various regression models in regions (here voivodeships). Numerous features of the proposed method require further investigation.

The proposed method provided good results for livestock activity data of agricultural sector. Apart from the reported above study, the approach was also applied in a residential sector for disaggregation of natural gas consumption in households. In that case, with disaggregation featured from voivodeships into municipalities, the results turned to be quite modest. This was partly due to limited spatial correlation of the analysed process and too large extent of disaggregation. The method is feasible for disaggregation from districts into municipalities, but not from voivodeships into municipalities.

It should be stressed that the primary asset of the proposed approach is the possibility to asses significance of considered regression coefficients. The widely used proportional distribution of activity data can be based ouly on expert judgements, providing no means for outcome verification.

Acknowledgement

The study was conducted within the 7FP Marie Curie Actions IRSES project No. 247645 *Geoinformation technologies, spatio-temporal approaches, and full carbon account for improving accuracy of GHG inventories*. The support from the Polish Ministry of Science and Higher Education within the funds for statutory works of young scientists is greatefully acknowledged.

This contribution is also supported by the Foundation for Polish Science under International PhD Projects in Intelligent Computing; project financed from The European Union within the Innovative Economy Operational Programme 2007-2013 and European Regional Development Fund.

This work was completed with the help of Olha Danylo and Rostyslav Bun from the Lviv Polytechnic National University, who provided comprehensive information on inventory data.

Bibliography

[1] Banerjee S., Carlin B.P., Gelfand A.E. (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall/CRC.

[2] Boychuk K., Bun R., Regional spatial cadastres of GHG emissions in Energy sector: Accounting for uncertainty, *Climatic Change*, under revision.

[3] 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(4):372-375.

[4] Cressie N.A.C. (1993) *Statistics for Spatial Data*, Wiley, New York.

[5] European Environment Agency (2000) Corine Land Cover 2000. `http://www.eea.europa.eu/data-and-maps/data` Accessed November 2012.

[6] Gelfand A.E., Diggle P.J., Fuentes M., Guttorp P., Eds. (2010) *Handbook of Spatial Statistics*, Chapman & Hall/CRC.

[7] Gotway C.A., Young L.J. (2002) Combining incompatible spatial data, *Journal of the American Statistical Association* 97: 632-648.

[8] Główny Urząd Statystyczny (2012) Bank Danych Lokalnych. `http://www.stat.gov.pl/bdlen/app/strona.html?p_name=indeks` Accessed November 2012.

[9] 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(1-2):263-276.

[10] Horabik J., Nahorski Z., Improving resolution of a spatial air pollution inventory with a statistical inference approach, *Climatic Change*, under revision.

[11] IPCC (1996) IPCC Guidelines for National Greenhouse Gas Inventories. Volume 1, 2, and 3. Intergovernmental Panel on Climate Change, London.

[12] 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.

[13] Lim B., Boileau P., Bonduki Y. et al. (1999) Improving the quality of national greenhouse gas inventories, *Environmental Science & Policy* 2: 335-346.

[14] Rypdal K., Winiwarter W. (2001) Uncertainties in greenhouse gas emission inventories - evaluation, comparability and implications, *Environmental Sciencc & Policy* 4:107-116.

[15] McMillan A.S., Holland D.M., Morara M., Fend J. (2010) Combining numerical model output and particulate data using Bayesian space-time modeling, *Environmetrics* 21:48-65.

Appendix

Table 5.1: List of voivodships

l	Voivodship
1	Dolnośląskie
2	Kujawsko-Pomorskie
3	Lubelskie
4	Lubuskie
5	Łódzkie
6	Małopolskie
7	Mazowieckie
8	Opolskie
9	Podkarpackie
10	Podlaskie
11	Pomorskie
12	Śląskie
13	Swiętokrzyskie
14	Warmińsko-Mazurskie
15	Wielkopolskie
16	Zachodniopomorskie