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 2

# The data set

Considered is a livestock dataset (cattle, pigs, horses, poultry, etc.) based on agricultural census 2010, and available from the Central Statistical Office of Poland - Local Data Bank [8]. The goal is to allocate relevant livestock amounts from district (powiaty) into municipality (gminy) levels.

In particular, for horses the data are available also in municipalities, and this fact enables verification of the proposed disaggregation method. Therefore, in what follows we consider the task of disaggregation of number of horses reported for 314 districts into 2171 municipalities, taking advantage of covariate information observable for municipalities, compare Figure 2.1. Only rural municipalities are considered in the study, see Figure 2.2.

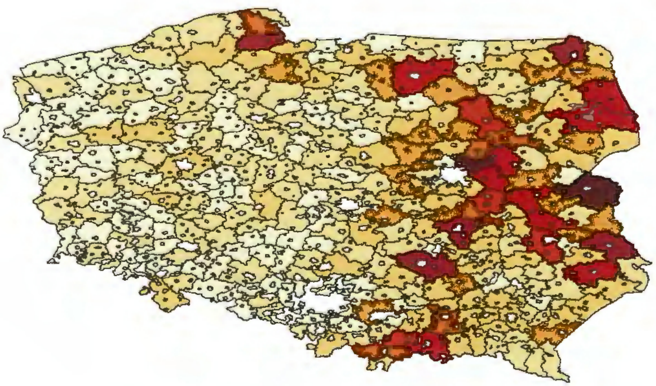

Figure 2.1: Livestock data (horses) available for districts

Figure 2.2: The net of rural municipalities

As explanatory variables we use population density (denoted $\boldsymbol{x}_1$) and land use information. For the latter, the CORINE Land Cover map, available from the European Environment Agency [5], was employed. For each rural municipality we calculate the area of agricultural classes, which may be related to livestock farming, see Figure 2.3. Three CORINE classes were considered (the CORINE class numbers are given in brackets):

- Arable land (2.1); denoted $\boldsymbol{x}_2$
- Pastures (2.3); denoted $\boldsymbol{x}_3$
- Heterogeneous agricultural areas (2.4), which include subclasses Complex cultivation patterns (2.4.2) and Land principally occupied by agriculture, with significant areas of natural vegetation (2.4.3); denoted $\boldsymbol{x}_4$.

The results of the disaggregation with the proposed procedure are further compared with the results of allocation proportional to population of municipalities. Here, we stress once more that only rural municipalities are considered in the study. Otherwise, allocation of number of horses in a proportion to population would be meaningless. This naive approach, however, gave rise for a modification of the basic version of the method. Namely, we account for the fact that a relationship of farmed livestock with available covariates is diversified across the country - we allow for various regression models for regions. In this particular case study, we treat 16 voivodeships (województwa) as regions. This extention of the model is described in section 3.2.

Figure 2.3: CORINE land use map of Poland, with the net of rural municipalities

# 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 |