

Uncertainty of Spatial Disaggregation Procedures: Conditional Autoregressive Versus Geostatistical Models

Joanna Horabik-Pyzel
 Systems Research Institute
 of Polish Academy of Sciences
 ul. Newelska 6, 01-447 Warsaw, Poland
 Email: joanna.horabik@ibspan.waw.pl

Zbigniew Nahorski
 Systems Research Institute PAS
 and Warsaw School of Information Technology
 ul. Newelska 6, 01-447 Warsaw, Poland
 Email: zbigniew.nahorski@ibspan.waw.pl

Abstract—Consider the problem of allocation of spatially correlated gridded data to finer spatial scale, conditionally on covariate information observable in a fine grid. Spatial dependence of the process can be captured with the conditional autoregressive structure, suitable for gridded (areal level) data. Also geostatistical methods, particularly empirical universal kriging, can be used for this purpose. In this study, we compare prediction results as well as prediction standard errors for two disaggregation procedures, based on the inventory of agricultural ammonia emissions reported in Pomeranian Voivodeship of Poland.

I. INTRODUCTION

IN MANY environmental and epidemiological applications, one has to deal with spatial variables observed at different resolutions. The change of support problem is encountered, for example, in a development of high-resolution inventories of greenhouse gases [1], [2] or ammonia emissions [3].

The choice of a relevant model capturing spatial correlation depends on a type of data, but also it can depend on a size of dataset and computational efficiency. In principle, the model suitable for areal data is based on Markov random fields, in particular the commonly used conditional autoregressive structure. However, the point-referenced data can be aggregated to the area level, and modelled the same way [4]. On the other hand, the geostatistical approach, designed for continuous spatial processes, can be used to model a process over a gridded domain.

In this paper, we aim to explore uncertainty underlying a particular procedure of areal data disaggregation from a coarse to fine grid. The setting assumes knowledge on (i) a variable of interest in a *coarse* grid, and (ii) some related variables (proxy data) in a *fine* grid. The task is accomplished with two alternative approaches to modelling spatial data: the one based on conditional autoregressive model, and the other, using geostatistical methods. These models represent two general classes commonly used in spatial statistics. Within geostatistical approach, empirical universal kriging was applied for a prediction of unknown values in a fine grid. For each model we compare the prediction standard errors against actual residuals (their absolute values), as resulting from an empirical study of

ammonia emission inventory. In addition, we analyse the effect of the level of disaggregation.

II. MOTIVATING DATA SET

A. Inventory of ammonia emissions

The analysed dataset concerns ammonia (NH_3) emission inventory in a region of Poland. Ammonia is emitted mainly (up to 80 – 90%) by agricultural sources such as livestock production and fertilized fields [5], [6]. High concentrations of ammonia can lead to acidification of soils [7], forest decline, and eutrophication of waterways [6]. All of these lead to loss of plant biodiversity [8]. Moreover, ammonia emissions are recognized for their importance in contributing to fine particulate matter [9], hence their spatial distribution is of great importance.

However, agricultural emission sources cannot be measured directly, and spatial emission patterns need to be assessed otherwise. This issue was addressed, for instance in [3], where agricultural and land cover data were used to disaggregate the national NH_3 emission totals across Great Britain. This was accomplished employing a spatially weighted redistribution of emission sources, with weights based on respective landcover classes. It was demonstrated in [10] that this type of straightforward, linear approaches to spatial allocation can be substantially improved by introducing a spatial random effect modelled with a conditional autoregressive structure.

B. Data description

The dataset comprises the gridded inventory of ammonia emissions from fertilization (in tonnes per year), reported in Pomeranian Voivodeship of Poland. The inventory grid cells are of regular $5\text{km} \times 5\text{km}$ size, and the whole of cadastral survey compiles $n = 800$ cells, denoted $\mathbf{y} = (y_1, \dots, y_n)^T$; see Fig. 1.

It should be noted, that the considered variable y of ammonia refers to a total amount of emissions over a grid cell; it is called an *extensive* variable [11]. This should be distinguished from *intensive* variables, e.g. emission concentrations or proportions over a geographic region.

For explanatory information we use the CORINE Land Cover Map for this region, available from the European Environment Agency [12]. Specifically, for each single grid cell we calculate area (in m^2) of those land use classes that are related to ammonia emissions. The following CORINE classes were considered (for reference, the CORINE class numbers are given in brackets):

- Non-irrigated arable land (211), denoted x_1 ;
- Fruit tree and berry plantations (222), denoted x_2 ;
- Pastures (231), denoted x_3 ;
- Complex cultivation patterns (242), denoted x_4 ;
- Principally agriculture, with natural vegetation (243), denoted x_5 .

Only land use data are used as explanatory information. Also, it should be pointed out that our modelling approach includes both a regression component as well as a spatial correlation component, and the resulting regression coefficients are not the same as typical emission coefficients, specific for the listed land use classes.

Performance of a disaggregation framework depends on various factors. Among others, it highly depends on the extent of disaggregation; note, that this is connected with a preservation of the correlation across spatial scales. An impact of this feature is evaluated in the study. We test the disaggregation from $10km \times 10km$ and $15km \times 15km$ (coarse) grids to a $5km \times 5km$ (fine) grid. To examine performance of the disaggregation procedure, first we aggregate the original fine grid emissions into respective coarse grid cells. Next, we fit respective model and predict ammonia emissions for a 5km fine grid. Finally, we check obtained results with the original inventory emissions of a 5km grid. Thus, our simulation study tests the cases of a fourfold and ninefold disaggregation. The aggregated values of the two coarse grids as well as the actual inventory data in the fine grid are shown in Fig. 1.

III. DISAGGREGATION MODEL BASED ON CONDITIONAL AUTOREGRESSIVE STRUCTURE

In this section we present an approach for areal to areal data realignment, where the residual covariance structure is modelled with the conditional autoregressive (CAR) specification [13], [14]. This class of models is used in the case of areal data, and it introduces spatial association through a neighbourhood structure.

A. The model

1) *Fine grid*: We begin with the model specification in a fine grid. Let $\mathbf{Y} = \{Y_i\}_{i=1}^n$ denote random variables associated with missing values of interest $\mathbf{y} = \{y_i\}_{i=1}^n$ defined at each cell i , $i = 1, \dots, n$ of a fine grid. Assume that random variables Y_i follow a Gaussian distribution with respective mean and variance, $Y_i | \mu_i, \sigma_Y^2 \sim \mathcal{N}(\mu_i, \sigma_Y^2)$. 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 underlying mean process, and the (missing) observations in a fine grid are related to this process through a measurement error of variance σ_Y^2 . The model for the underlying mean process is formulated

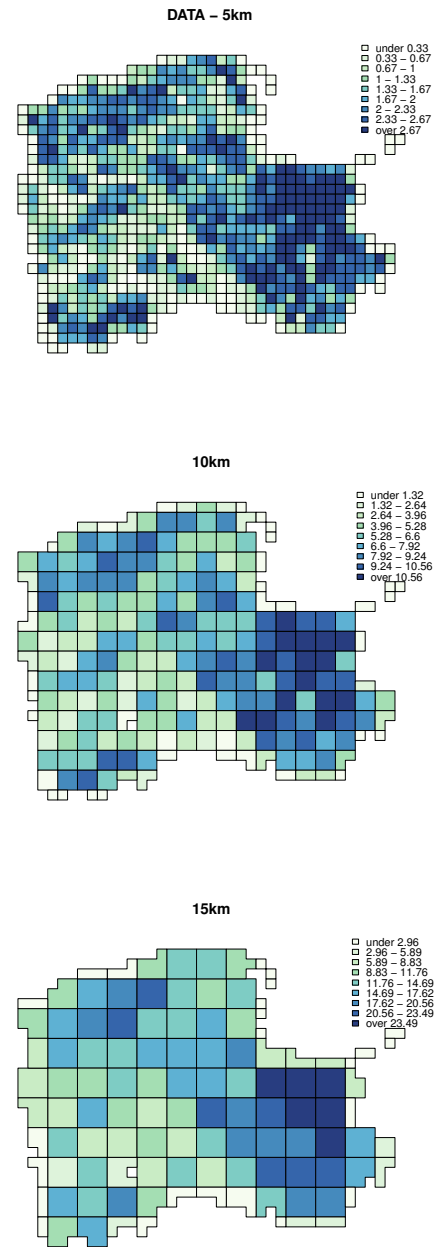


Fig. 1. Ammonia emissions (in tonnes/year): inventory data in 5km grid, and aggregated values in 10km and 15km grids

as a sum of regression component with available covariates, and a spatially varying random effect.

The approach to modeling 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 conditional autoregressive CAR model. 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 $i = 1, \dots, n$ [15], [16]

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

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 ; $\mathbf{x}_i^T \boldsymbol{\beta}$ is a regression component with explanatory covariates for area i and a respective vector of regression coefficients.; and τ^2 is a variance parameter.

The conditionals (1) yield the following joint distribution of the process $\boldsymbol{\mu}$, see e.g. [15]

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

where \mathbf{X} is a design matrix with vectors \mathbf{x}_i ; \mathbf{D} is an $n \times n$ diagonal matrix with w_{i+} on the diagonal; and \mathbf{W} is an $n \times n$ matrix with adjacency weights w_{ij} . Equivalently, we can write (2) as

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{N}), \quad (3)$$

denoting also $\mathbf{N} = \tau^2 (\mathbf{D} - \rho \mathbf{W})^{-1}$.

2) *Coarse grid*: The model for a coarse grid (aggregated) observed data is obtained by multiplication of the mean process (3) with an $N \times n$ aggregation matrix \mathbf{C} , where N is a number of observations in a coarse grid

$$\mathbf{C}\boldsymbol{\mu} = \mathbf{C}\mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\epsilon}, \quad \mathbf{C}\boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \mathbf{C}\mathbf{N}\mathbf{C}^T). \quad (4)$$

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

$$\mathbf{Z} | \boldsymbol{\lambda} \sim \mathcal{N}_N(\boldsymbol{\lambda}, \sigma_Z^2 \mathbf{I}_N), \quad (5)$$

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

B. Maximum likelihood estimation

The parameters $\boldsymbol{\beta}$, σ_Z^2 , τ^2 and ρ are estimated with the maximum likelihood method based on the joint unconditional distribution of \mathbf{Z}

$$\mathbf{Z} \sim \mathcal{N}_N(\mathbf{C}\mathbf{X}\boldsymbol{\beta}, \mathbf{M} + \mathbf{C}\mathbf{N}\mathbf{C}^T), \quad (6)$$

where $\mathbf{M} = \sigma_Z^2 \mathbf{I}_N$. Next, the log likelihood function associated with (6) is formulated

$$L(\cdot) = -\frac{1}{2} \log |\mathbf{M} + \mathbf{C}\mathbf{N}\mathbf{C}^T| - \frac{N}{2} \log(2\pi) - \frac{1}{2} (\mathbf{z} - \mathbf{C}\mathbf{X}\boldsymbol{\beta})^T (\mathbf{M} + \mathbf{C}\mathbf{N}\mathbf{C}^T)^{-1} (\mathbf{z} - \mathbf{C}\mathbf{X}\boldsymbol{\beta}),$$

where $|\cdot|$ denotes the determinant. 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 parameter estimators for this model have been developed by means of the expected and observed Fisher information matrices, details of which are provided in [17].

C. Prediction in a fine grid

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

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

The above distribution allows for full inference, yielding both the predictor

$$E(\widehat{\boldsymbol{\mu}} | \mathbf{z}) = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{N}}\mathbf{C}^T (\widehat{\mathbf{M}} + \mathbf{C}\widehat{\mathbf{N}}\mathbf{C}^T)^{-1} [\mathbf{z} - \mathbf{C}\mathbf{X}\widehat{\boldsymbol{\beta}}] \quad (7)$$

and its variance

$$Var(\widehat{\boldsymbol{\mu}} | \mathbf{z}) = \widehat{\mathbf{N}} - \widehat{\mathbf{N}}\mathbf{C}^T (\widehat{\mathbf{M}} + \mathbf{C}\widehat{\mathbf{N}}\mathbf{C}^T)^{-1} \mathbf{C}\widehat{\mathbf{N}}. \quad (8)$$

IV. GEOSTATISTICAL APPROACH

In this section, we briefly review geostatistical approach, which is dedicated to modelling point-referenced data over a continuous domain. It specifies the process through a covariance function.

In the application considered, ammonia emission $Y(\mathbf{s})$ is the variable of area type. The point level data are obtained by dividing this variable by area (in km²) of respective grid cell. This is a kind of approximation which expresses emissions over a unit area, or (roughly) *emission intensity*. Thus, a geostatistical model is applied to the modified process $Y_A(\mathbf{s}) = Y(\mathbf{s})/A$, where A stands for area of a 5km grid cell. We observe $\mathbf{Y}_A = (Y_A(\mathbf{s}_1), \dots, Y_A(\mathbf{s}_n))^T$ in a fine grid, and we wish to predict the variable $Y_A(\mathbf{s}_0)$ at a location $\mathbf{s}_0 \in D$ where it has not been observed, i.e. in centroids of a coarse grid.

Gaussian geostatistical models are based on two assumptions: second order stationarity and isotropy. Second order stationarity means that the process mean is constant and its covariance function depends only on the difference between locations. The process is isotropic if, additionally, the covariance depends only on distance (not direction) between two locations. Once these assumptions are met, spatial process can be modelled with parametric covariance functions. The exponential covariance function, applied in this study, is defined as,

$$cov(h) = \begin{cases} \sigma^2 \exp(-\phi h) & \text{if } h > 0 \\ \tau_{nug}^2 + \sigma^2 & \text{if } h = 0, \end{cases} \quad (9)$$

where h denotes the Euclidean distance between two points, τ_{nug}^2 represents the nugget effect, σ^2 is the partial sill, and ϕ denotes the effective range of the covariance. Furthermore, denote $\mathbf{K} = cov(\mathbf{Y}_A, \mathbf{Y}_A^T)$, $\mathbf{k} = cov\{Y_A, Y_A(\mathbf{s}_0)\}$, and $k_0 = cov\{Y_A(\mathbf{s}_0), Y_A(\mathbf{s}_0)\}$.

With land use information available as covariates denoted $\mathbf{X} = (\mathbf{x}(s_1), \dots, \mathbf{x}(s_n))^T$ and $\mathbf{x}(s_0)$, universal kriging [18] was applied for a prediction in a fine grid. The model for a random field $Y_A(s)$ has a linear mean function and the error process $\epsilon(s)$

$$Y_A(s) = \mathbf{x}(s)^T \boldsymbol{\beta} + \epsilon(s), \quad (10)$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is a vector of p unknown coefficients, and $\epsilon(s)$ is a zero-mean Gaussian process with the exponential covariance function given by (9).

The geostatistical prediction problem is formulated as follows. We seek a predictor $\hat{Y}_A(s_0)$ that minimises the mean squared prediction error among the predictors satisfying two properties:

- 1) linearity: $\hat{Y}_A(s_0) = \boldsymbol{\lambda}^T \mathbf{Y}_A$
- 2) unbiasedness: $E\hat{Y}_A(s_0) = E(\boldsymbol{\lambda}^T \mathbf{Y}_A) = EY_A(s_0)$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$

We obtain minimisation of $E\{Y_A(s_0) - \boldsymbol{\lambda}^T \mathbf{Y}_A\}^2$ subject to

$$\mathbf{X}^T \boldsymbol{\lambda} = \mathbf{x}(s_0). \quad (11)$$

This constraint optimisation task can be solved with the method of Lagrange multipliers, see e.g. [19], [15]. Provided that matrices \mathbf{K} and $\mathbf{X}^T \mathbf{K}^{-1} \mathbf{X}$ are invertible, it yields

$$\boldsymbol{\lambda} = \left\{ \mathbf{K}^{-1} - \mathbf{K}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{K}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K}^{-1} \right\} \mathbf{k} + \mathbf{K}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{K}^{-1} \mathbf{X})^{-1} \mathbf{x}(s_0) \quad (12)$$

and the best linear unbiased predictor (BLUP) of $Y_A(s_0)$ becomes

$$\begin{aligned} \hat{Y}_A(s_0) &= \boldsymbol{\lambda}^T \mathbf{Y}_A \\ &= \left\{ \mathbf{k} + \mathbf{X} (\mathbf{X}^T \mathbf{K}^{-1} \mathbf{X})^{-1} [\mathbf{x}(s_0) - \mathbf{X}^T \mathbf{K}^{-1} \mathbf{k}] \right\}^T \\ &\quad \times \mathbf{K}^{-1} \mathbf{Y}_A. \end{aligned} \quad (13)$$

The resulting mean squared error of the BLUP, called also the kriging variance, is given by

$$k_0 - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} + \boldsymbol{\gamma}^T (\mathbf{X}^T \mathbf{K}^{-1} \mathbf{X})^{-1} \boldsymbol{\gamma}, \quad (14)$$

where $\boldsymbol{\gamma} = \mathbf{x}(s_0) - \mathbf{X}^T \mathbf{K}^{-1} \mathbf{k}$.

Estimation of parameters has been performed using the `geoR` package from the R software [20].

V. RESULTS

A. Fourfold disaggregation

This subsection presents the model testing results for disaggregation from a 10km grid.

Table I displays the maximum likelihood estimates and standard errors for all parameters. Also the statistical significance of regression coefficients is reported with the t-statistic and respective p -values. It should be stressed, that estimation of parameters has been performed for *emission values* in the case of CAR models, and for *emission intensity* in the

case of geostatistical models (denoted GEOST). Therefore, parameter estimates are comparable only within the same class of models. As regards GEOST models, due to an optimisation procedure [20], the standard errors are available only for the ratio $\tau_{nug}^2 / \sigma_Z^2$.

From a visual comparison of the 5km maps with predicted values of ammonia emissions (not shown), the differences with respect to the original data cannot be easily distinguished. Instead, Fig. 2 presents scatterplots of predicted values y_i^* against observations y_i . This suggests that CAR model gives better results than GEOST. In general, CAR model provides very accurate predictions, although it tends to overestimate significantly some of small values.

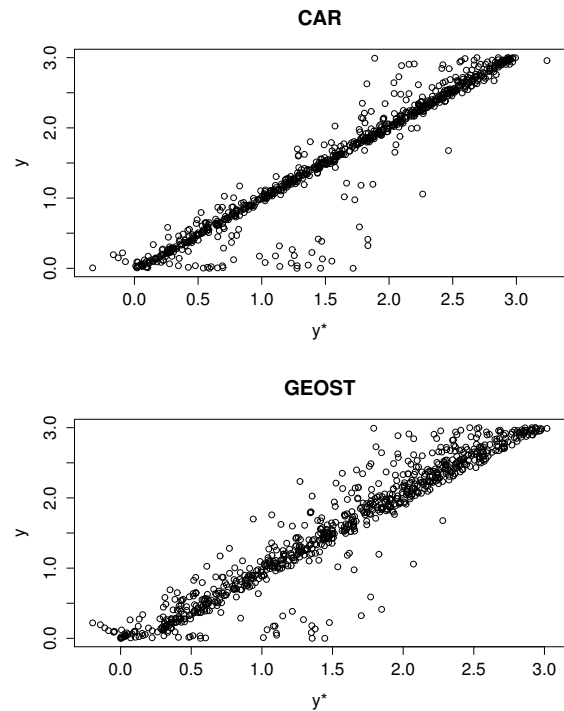


Fig. 2. Predicted (y^*) versus observed (y) values (in tonnes/year); disaggregation from 10km grid

Model residuals ($d_i = y_i - y_i^*$) are further summarised in Table II (the upper panel). The quantitative comparison confirms that CAR model outperforms the geostatistical one, both in terms of the mean squared error (mse) as well as the sample correlation coefficient (r). Still, the highest overestimate, i.e. $\min(d_i) = -1.717$, is reported for CAR model.

Next, the prediction standard error was calculated following the formula (8) for CAR model, and the formula (14) for geostatistical model. Since in the present case study the correct values of emissions predicted in 5km grid are known, we are in a position to compare the prediction error with actual residuals (more precisely, with their absolute values). In Fig. 3 these values are presented on maps for both disaggregation procedures. For CAR model, it is noticeable that the prediction error does not reflect diversification of actual residuals very ac-

TABLE I
MAXIMUM LIKELIHOOD ESTIMATES FOR COARSE GRIDS

	Est.	Std.Err.	t-statistic	p-value	Est.	Std.Err.	t-statistic	p-value
CAR models								
	10km				15km			
β_1	1.13e-07	3.26e-09	34.66	2.99e-59	1.12e-07	3.95e-09	28.26	6.37e-51
β_2	2.56e-07	1.94e-07	1.31	0.09	-	-	-	-
β_3	9.77e-08	1.19e-08	8.20	3.34e-13	1.07e-07	1.84e-08	5.83	3.11e-08
β_4	1.18e-07	2.13e-08	5.51	1.27e-07	1.24e-07	2.77e-08	4.49	9.02e-06
β_5	1.27e-07	1.32e-08	9.57	2.92e-16	1.27e-07	1.74e-08	7.31	2.84e-11
σ_Z^2	0.334	0.073	-	-	2.339	0.424	-	-
τ^2	0.536	0.082	-	-	0.214	0.088	-	-
ρ	0.948	9.98e-04	-	-	0.966	4.91e-04	-	-
GEOST models								
	10km				15km			
β_1	9.72e-08	5.83e-09	16.68	1.80e-31	9.21e-08	8.75e-09	10.53	2.19e-18
β_2	-	-	-	-	-	-	-	-
β_3	8.06e-08	1.62e-08	4.96	1.36e-06	-	-	-	-
β_4	9.53e-08	3.65e-08	2.61	0.005	1.21e-07	5.69e-08	2.12	0.018
β_5	1.12e-07	2.30e-08	4.88	1.91e-06	1.12e-07	3.79e-08	2.96	0.001
σ_Z^2	2.04e-03	-	-	-	4.50e-04	-	-	-
τ_{nug}^2	9.92e-05	0.07	-	-	9.84e-05	0.285	-	-
ϕ	205.01	298.41	-	-	61.02	61.39	-	-

TABLE II
ANALYSIS OF RESIDUALS

	mse	min(d_i)	max(d_i)	r
10km grid				
CAR	0.064	-1.717	1.104	0.961
GEOST	0.077	-1.444	1.200	0.956
15km grid				
CAR	0.136	-2.428	0.646	0.915
GEOST	0.144	-1.914	1.519	0.913

curately, and the highest values of residuals are underestimated (compare the scales of both maps). Otherwise, the prediction standard errors seem to provide a reasonable assessment. On the other hand, the prediction standard error for GEOST model is significantly underestimated, as seen from the map scales. Note, that for both disaggregation methods, the highest residuals are reported on the border of the domain; this fact is known in spatial statistics as the *edge effect*.

In addition, Fig. 4(a) presents the differences between the prediction standard errors of the models and absolute values of actual residuals. The empirical cumulative distributions of these differences confirm that the geostatistical model underestimates the prediction standard errors, much more than the CAR model does.

B. Ninefold disaggregation

Next, the results of disaggregation from a 15km grid are presented. In Table I we can see, for respective models, the increase of variances σ_Z^2 when turning from a 10km to 15km disaggregation.

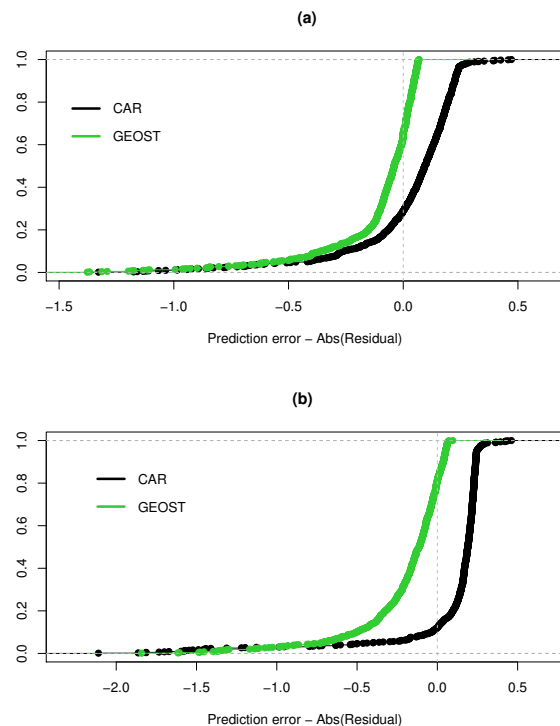


Fig. 4. Empirical cumulative distributions of the differences between the prediction standard errors and absolute values of residuals (in tonnes/year) for (a) 10km and (b) 15km disaggregation

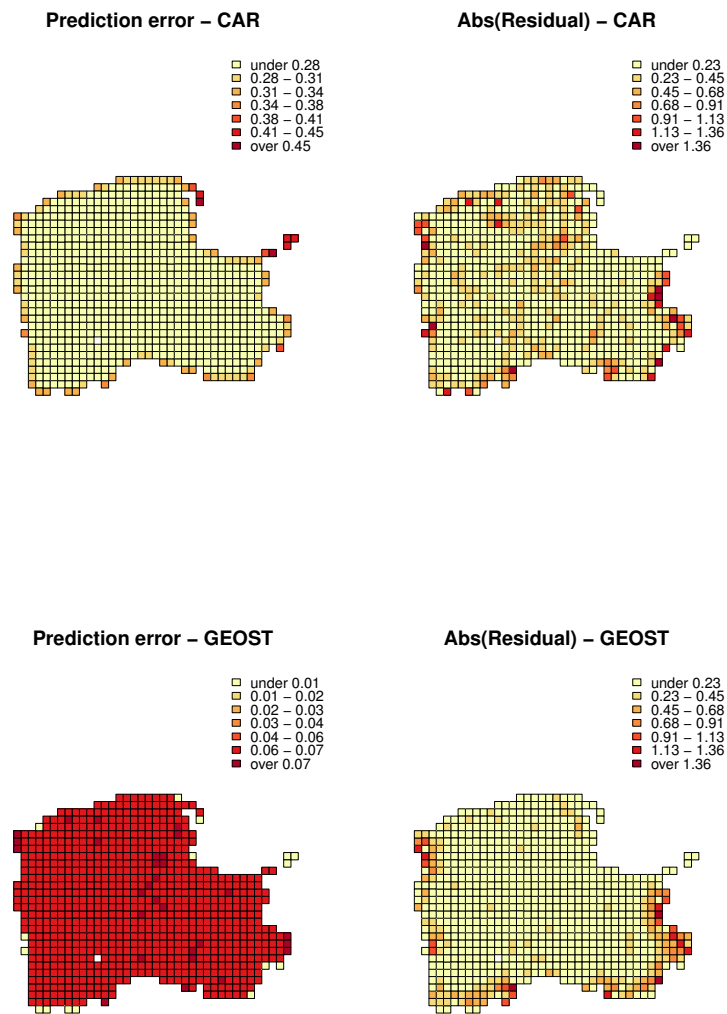


Fig. 3. Prediction standard errors and absolute values of residuals (in tonnes/year) for CAR (upper panel) and GEOST (lower panel) models; disaggregation from 10km grid. Note that the maps are drawn in different scales.

The scatterplot in Fig. 5 reveals important differences between the two methods. CAR model generally provides more accurate predictions but heavily overestimates numerous values. Predictions from GEOST are less accurate, and the model also tends to overestimate low values and underestimate high ones. Overall, Table II shows that both approaches provide comparable quality of predictions, as summarised by the mean squared error and correlation coefficient. Again, the highest overestimate, i.e. $\min(d_i) = -2.428$, is noted for CAR model.

For the case of ninefold disaggregation, the prediction standard errors and absolute values of residuals are depicted

in Fig. 6. For CAR model, this comparison provides quite good picture, although the highest values of residuals are still underestimated. Apart from this, the model uncertainty is reflected rather well. This is not the case for GEOST model. Firstly, a regular pattern on the map of prediction standard error is completely different from the actual residuals. It can be attributed to inherent features of the geostatistical method which provides the lowest prediction error at observed locations. Secondly, the prediction error for this procedure is evidently underestimated, similarly as for the case of 10km disaggregation.

Respective cumulative distributions of the differences be-

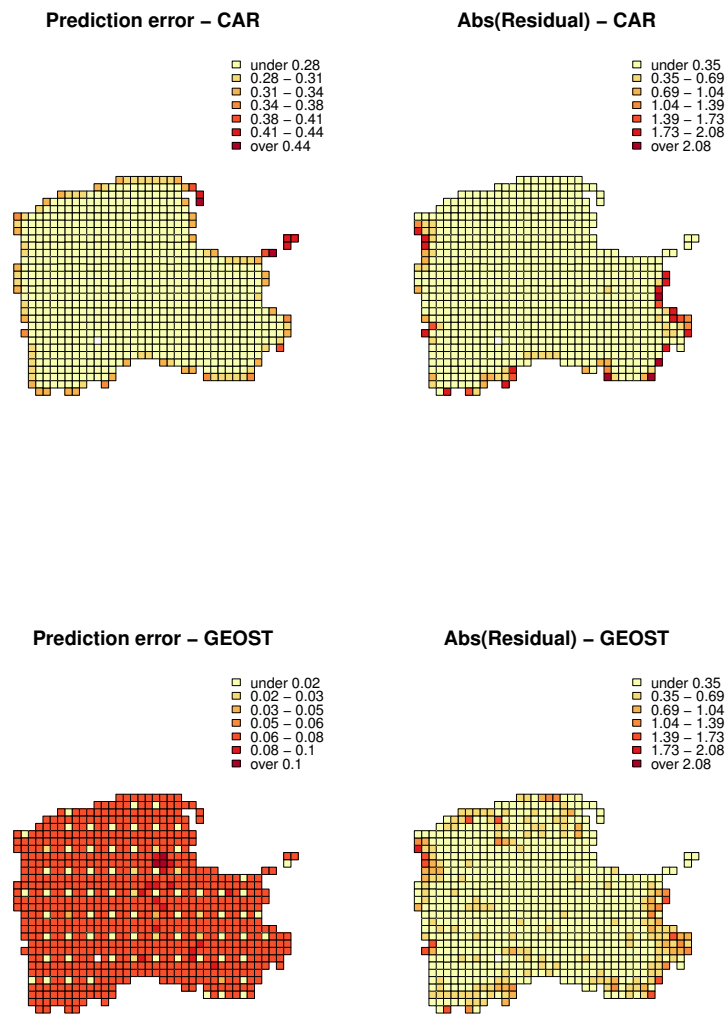


Fig. 6. Prediction standard errors and absolute values of residuals (in tonnes/year) for CAR (upper panel) and GEOST (lower panel) models; disaggregation from 15km grid. Note that the maps are drawn in different scales.

tween the prediction standard errors and absolute values of residuals, presented in Fig. 4(b), illustrate that CAR model provides rather higher estimates of error than the actual residuals, while GEOST underestimates the model error. Compared with the results for 10km disaggregation, we note that accuracy of uncertainty assessment improved for CAR model, and it declined for GEOST.

VI. CONCLUDING REMARKS

The major objective of this paper was to study uncertainty of two procedures for spatial allocation from a coarse to fine grid. For a particular disaggregation setting with proxy data

available in a fine grid, we analysed the approach based on the conditional autoregressive structure, and the one based on the geostatistical methods.

For disaggregations from 10km and 15km grids to a 5km grid, both methods provided very good predictions, with $r = 0.96$ and 0.91 , respectively. The mean squared error of predictions was approximately 20% lower for CAR model in the case of fourfold disaggregation, and 5% in the case of ninefold disaggregation.

As regards the geostatistical method, despite its good predictive performance, this disaggregation procedure failed to properly assess uncertainty of the model. It should be noted

TABLE III
PROS AND CONS OF THE DISAGGREGATION METHODS

CAR model	GEOST model
<i>ADVANTAGES</i>	
<ul style="list-style-type: none"> - Very good predictive performance - Reliable assessment of prediction error - Accuracy of uncertainty assessment remains high also when increasing a degree of disaggregation. - The method is well suited for areal data. 	<ul style="list-style-type: none"> - Very good predictive performance - Prediction with universal kriging is a well known, popular method, and thus easy to implement for practitioners. - Wide availability of dedicated software
<i>DISADVANTAGES</i>	
<ul style="list-style-type: none"> - CAR structure is less popular among practitioners, and usually one needs to develop their own codes. 	<ul style="list-style-type: none"> - Poor assessment of model uncertainty - The method is dedicated to point-referenced data, and application for areal data requires some additional manipulations.

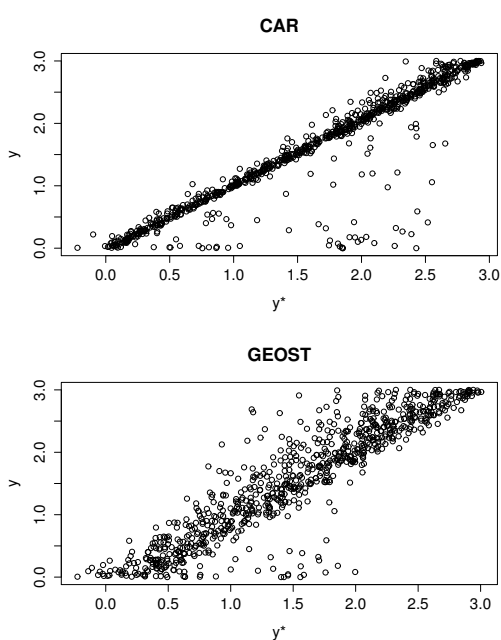


Fig. 5. Predicted (y^*) versus observed (y) values (in tonnes/year); disaggregation from 15km grid

that in our study the covariance function was unknown and respective parameters were estimated, which resulted in *empirical* universal kriging procedure. In such a case, the predictor is no longer a linear function of the data. In [16] the authors note that empirical kriging variance tends to underestimate the actual prediction error variance of the empirical universal kriging predictor because it does not account for additional error due to parameter estimation. CAR model provided a reliable assessment of prediction error. In this particular case study, this might result also from the fact that CAR structure is dedicated to areal data, like the analysed dataset of ammonia emissions.

When increasing a degree of disaggregation, obviously the quality of predictions decreases, but the accuracy of uncertainty assessment improved for CAR model. For the geosta-

tistical approach presented, it is generally poor. Nevertheless, universal kriging is a popular, widespread technique, built into numerous software tools, which facilitates application of this disaggregation approach.

To summarise, Table III lists advantages and disadvantages of both methods for the considered case of areal data disaggregation.

ACKNOWLEDGMENT

The authors gratefully acknowledge the staff of *Ekometria sp. z o.o.*, in particular Wojciech Trapp and Małgorzata Paciorek, for provision of data.

REFERENCES

- [1] K. Boychuk and R. Bun, "Regional spatial inventories (cadastres) of GHG emissions in the Energy sector: Accounting for uncertainty," *Climatic Change*, vol. 124, pp. 561–574, 2014. doi: 10.1007/s10584-013-1040-9. [Online]. Available: <http://dx.doi.org/10.1007/s10584-013-1040-9>
- [2] R. Bun, K. Hamal, M. Gusti, and A. Bun, "Spatial GHG inventory at the regional level: accounting for uncertainty," *Climatic Change*, vol. 103, no. 1-2, pp. 227–244, 2010. doi: 10.1007/s10584-010-9907-5. [Online]. Available: <http://dx.doi.org/10.1007/s10584-010-9907-5>
- [3] U. Dragosits, M. Sutton, C. Place, and A. Bayley, "Modelling the spatial distribution of agricultural ammonia emissions in the UK," *Environ. Pollut.*, vol. 102, no. S1, pp. 195–203, 1998. doi: 10.1016/S0269-7491(98)80033-X. [Online]. Available: [http://dx.doi.org/10.1016/S0269-7491\(98\)80033-X](http://dx.doi.org/10.1016/S0269-7491(98)80033-X)
- [4] H. Song, M. Fuentes, and S. Ghosh, "A comparative study of Gaussian geostatistical models and Gaussian Markov random field models," *J. Multivariate Anal.*, vol. 99, pp. 1681–1697, 2008. doi: 10.1016/j.jmva.2008.01.012. [Online]. Available: <http://dx.doi.org/10.1016/j.jmva.2008.01.012>
- [5] T. Misselbrook, T. Van Der Weerden, B. Pain, S. Jarvis, B. Chambers, K. Smith, V. Phillips, and T. Demmers, "Ammonia emission factors for UK agriculture," *Atmos. Environ.*, vol. 34, pp. 871–880, 2000. doi: 10.1016/S1352-2310(99)00350-7. [Online]. Available: [http://dx.doi.org/10.1016/S1352-2310\(99\)00350-7](http://dx.doi.org/10.1016/S1352-2310(99)00350-7)
- [6] G. Velthof, C. van Bruggenb, C. Groenesteinc, B. de Haand, M. Hoogeveene, and J. Huijsmans, "A model for inventory of ammonia emissions from agriculture in the Netherlands," *Atmos. Environ.*, vol. 46, pp. 248–255, 2012. doi: 10.1016/j.atmosenv.2011.09.075. [Online]. Available: <http://dx.doi.org/10.1016/j.atmosenv.2011.09.075>
- [7] P. Barak, B. Jobe, A. Krueger, L. Peterson, and D. Laird, "Effects of long-term soil acidification due to nitrogen fertilizer inputs in Wisconsin," *Plant Soil*, vol. 197, pp. 61–69, 1997. doi: 10.1023/A:1004297607070. [Online]. Available: <http://dx.doi.org/10.1023/A:1004297607070>

- [8] R. Bobbink, M. Hornung, and J. Roelofs, "The effects of airborne nitrogen pollutants on species diversity in natural and semi-natural European vegetation," *J. Ecol.*, vol. 86, pp. 717–738, 1998. doi: 10.1046/j.1365-2745.1998.8650717.x. [Online]. Available: <http://dx.doi.org/10.1046/j.1365-2745.1998.8650717.x>
- [9] J. Erisman and M. Schaap, "The need for ammonia abatement with respect to secondary PM reductions in Europe," *Environ. Pollut.*, vol. 129, pp. 159–163, 2004. doi: 10.1016/j.envpol.2003.08.042. [Online]. Available: <http://dx.doi.org/10.1016/j.envpol.2003.08.042>
- [10] J. Horabik and Z. Nahorski, "Improving resolution of a spatial inventory with a statistical inference approach," *Climatic Change*, vol. 124, no. 3, pp. 575–589, 2014. doi: 10.1007/s10584-013-1029-4. [Online]. Available: <http://dx.doi.org/10.1007/s10584-013-1029-4>
- [11] P. Legendre and L. Legendre, *Numerical Ecology*, ser. Developments in Environmental Modelling. Elsevier Science, 2012. ISBN 9780444538697
- [12] European Environment Agency, "Corine Land Cover 2000," <http://www.eea.europa.eu/data-and-maps/data>, 2000.
- [13] J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *J. Roy. Stat. Soc. B*, vol. 36, pp. 192–236, 1974. [Online]. Available: <http://www.jstor.org/stable/2984812>
- [14] —, "Statistical analysis of non-lattice data," *J. Roy. Stat. Soc. D-Stat.*, vol. 24, no. 3, pp. 179–195, 1975. doi: 10.2307/2987782. [Online]. Available: <http://dx.doi.org/10.2307/2987782>
- [15] N. Cressie, *Statistics for spatial data*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993. ISBN 9780471002550. [Online]. Available: <http://books.google.pl/books?id=4SdRAAAAMAAJ>
- [16] A. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of Spatial Statistics*, ser. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2010. ISBN 9781420072884
- [17] J. Horabik and Z. Nahorski, "The Cramer-Rao lower bound for the estimated parameters in a spatial disaggregation model for areal data," in *Intelligent Systems 2014*, P. Angelov, K. Atanassov, L. Doukowska, M. Hadjiski, V. Jotsov, and J. Kacprzyk, Eds. Springer International Publishing, 2015. doi: 10.1007/978-3-319-11313-5. ISBN 978-3-319-11312-8 pp. 661–668.
- [18] S. Banerjee, B. Carlin, and A. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, ser. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 2004. ISBN 9780203487808. [Online]. Available: <http://books.google.pl/books?id=YqpZKTp-Wh0C>
- [19] M. Stein, *Interpolation of Spatial Data. Some Theory for Kriging*, ser. Springer Series in Statistics. Springer, New York, 1999.
- [20] P. Ribeiro Jr. and P. Diggle, "geoR: a package for geostatistical analysis," *R-NEWS*, vol. 1, no. 2, pp. 15–18, 2001. [Online]. Available: <http://cran.R-project.org/doc/Rnews>