Bayesian small area models for investigating spatial heterogeneity and factors affecting the amount of solid waste in Italy

C. Calculli and S. Arima

Abstract This study aims at investigating factors that impact solid waste generation rates at the Italian province-level scale. An approach based on Bayesian small area models is used to evaluate potential spatial dependencies and local variations on the distribution of waste generation rates and its determinants. An extension of the basic Fay-Herriot model to include spatial correlation among neighboring areas is considered (SFH). Preliminary results suggest a highest per capita production of differentiated (or recyclable) waste fraction in wealthier Italian provinces highlighting the intrinsic correlation between areas characterized by similar socio-economic scenarios.
Abstract *L'obiettivo di questo lavoro è quello di analizzare i fattori che influenzano i tassi di produzione dei rifiuti solidi urbani su scala provinciale in Italia. Un approccio modellistico basato sulla stima bayesiana per piccole aree viene utilizzato per valutare la dipendenza spaziale e le variazioni terriroriali nella distribuzione dei tassi di produzione dei rifiuti e delle loro determinanti. In particolare, viene considerata una estensione del modello di Fay-Herriot (SFH) al caso di correlazione spaziale tra dati di aree adiacenti. I risultati preliminari suggeriscono una produzione maggiore pro-capite della frazione differenziata (o riciclabile) di rifiuti urbani nelle province italiane più ricche, evidenziando una correlazione intrinseca tra aree caratterizzate da scenari socio-economici simili.*

Key words: Bayesian SAE, Spatial random effects models, Municipal solid wastes

---

Crescenza Calculli
Department of Economics and Finance, University of Bari Aldo Moro, Largo Abbazia S. Scolastica - 70124 Bari, Italy e-mail: `crescenza.calculli@uniba.it`

Serena Arima
Department of History, Society and Human Studies, University of Salento, Piazza Tancredi, n.7 - 73100 Lecce, Italy e-mail: `serena.arima@unisalento.it`

1 Introduction

Municipal solid waste (MSW) commonly refers to the semi-solid or solid materials disposed by residents of urban areas. Due to population growth, urbanization, economic development, inappropriate recycling and governance programmes, MSW has become a global issue that poses serious threats to the environment and human health. Several industrialized countries have raised concerns about the economic viability and environmental acceptability of waste-disposal practices. For this reason, estimating waste production and treatment are crucial to quantify impacts, plan capacities and set policy targets. Several studies showed the impacts of socio-economic and demographic factors on solid waste generation trends, although neglecting the potential spatial dependency at local level. Few studies addressed the spatial characteristics of MSW generation rates applying spatial stratification or visually mapping their distributions using spatial statistical techniques such as simultaneous spatial autoregression (SAR) and geographically weighted regression (GWR) [7, 4]. An alternative model-based approach is proposed in order to capture the local variations in the distribution of waste generation rates accounting for their determinants at the Italian province level. This approach, based on methods of Small Area Estimation (SAE), assumes area-specific random effects to account for the between-area variation. A detailed overview of this methods is reported in Section 2.

2 Small area models

In recent years, small area estimation established as an important area of statistics as private and public agencies try to extract the maximum information from sample survey data. Small area methods arise when one wants to use sample surveys, generally designed to provide estimates of total or means for large subpopulations or domains, to draw inferences about smaller domains, such as as states, provinces, or different racial and/or ethnic subgroups that are unplanned by design. These small domains are called small areas. In recent years, demand for reliable estimates for small area estimates has greatly increased due to their growing use in formulating policies and programmes, allocating government funds, regional planning and other uses. Policy makers are often interested in targeting areas with particular needs in order to conduct specific actions: as in our case study, identifying those areas with highest waste production is of fundamental importance for decision makers who should plan the opening of new filtration and recycling plant or monitoring the consequences of such a waste production on the health of the population.
The simplest approach to small area estimation is to consider *direct estimators*, that is estimating the variable of interest using the domain-specific sample data. However, it is well known that the domain sample sizes are rarely large enough to support and accurate direct estimators [5]. Small area estimation tackles the problem of providing estimates of one or several variables of interest in areas where the infor-

mation available on those variables is, on its own, not sufficient to provide accurate direct estimates. Estimates for all areas are produced using the sample and some additional auxiliary information which should be available for all small areas. *Indirect estimators* are often employed in order to increase the effective domain sample size by borrowing strength from the related areas using linking models, census, administrative data and other auxiliary variables associated with the small areas. The model-based approach is widely used to develop indirect estimates. Depending on the type of data available, small area models are classified into two types, area-level and unit-level:

- area-level models where only area-level data are available for the response variable and the covariates;
- unit-level models where data on the response variable, and possibly on covariates, are available at the unit level.

The model proposed by [2] is the most popular small area model when data are available at area-level. It borrows strength from data available from all areas by assuming a hierarchical structure and uses auxiliary information from other data sources such as administrative records or censuses. The frequentist predictor of small area means, which is also known as empirical best linear unbiased predictor (EBLUP), results in a convex combination of the direct estimator and the synthetic estimator from the model. Properties of the predictors of small area means, such as bias and mean squared error, are derived conditionally on the auxiliary information. In this paper we focus on an area-level nested error linear regression model and propose a Bayesian hierarchical model for estimating finite population small area means.

2.1 The proposed model

Let m denote the number of small areas under consideration and let θ_i be the population characteristic of interest in the $i-$th area. The quantity of interest may be a total, a mean (or a proportion). Let y_i be a direct estimator of θ_i for area i and let X_i be the p dimensional vector of auxiliary data collected at the area level. The Fay-Herriot model is defined as

$$y_i = X_i'\beta + v_i + e_i, i = 1, \dots, m, \tag{1}$$

where the random effects $v_1, \dots, v_m$ and sampling errors $e_1, \dots, e_m$ are independent with $v_i \sim N(0, \sigma_v^2)$ and $e_i \sim N(0, \psi_i)$. We assume that the ψ_i's are known but, in general, different, reflecting the potential difference in sample sizes [5]. Our goal is the prediction of small area mean

$$\theta_i = X_i'\beta + v_i \tag{2}$$

The best linear predictor of θ_i for model (1) when model parameters (β, σ_v^2) are known is the EBLUP estimator $\tilde{\theta}_i = \gamma_{iv} y_i + (1-\gamma_{iv})X_i'\beta$, with $\gamma_{iv} = \sigma_v^2/(\sigma_v^2 + \psi_i)$.

In this paper, we consider a Bayesian extension of the Fay-Herriot model in Equation (1) and rewrite the model as the following multi-stage model:

Stage 1. $y_i = \theta_i + e_i \quad i = 1,\dots,m$, with $e_i \overset{\text{iid}}{\sim} N(0,\psi_i)$ and ψ_i a-priori known;
Stage 2. $\theta_i = x_i^T\beta + v_i \quad i = 1,\dots,m$, with $v_i \overset{\text{iid}}{\sim} N(0,\sigma_v^2)$;
Stage 3. β, σ_v^2 are, loosely speaking, mutually independent with flat priors over location parameters and inverse gamma distributions over the scale parameters.

In particular, we assume a Normal flat prior for the regression coefficients and an inverse gamma with parameters equal to 0.001 for the scale parameter σ_v^2.
Following [6], we extend the aforementioned model to incorporate spatially correlated random effects in the linking model: in particular, we add a spatial random effect denoted with b_i in the linking model in Stage 2 as follows:

$$\theta_i = x_i^T\beta + b_i \tag{3}$$

where

$$\mathbf{b} \sim MVN(\mathbf{0}, \Sigma(\sigma_b^2,\lambda))$$

and

$$\Sigma(\sigma_b^2,\lambda) = \sigma_b^2 \qquad\qquad D = \lambda\mathbf{R} + (1-\lambda)\mathbf{I}$$

defining a conditional auto-regressive (CAR) model on the area specific spatial effects. In particular, σ_b^2 is a spatial dispersion parameter and λ is a spatial autocorrelation parameter ($\lambda \in [0,1]$) and $\mathbf{I}$ is an identity matrix of dimension m. The matrix $\mathbf{R}$, commonly known as the neighbourhood matrix, has ith diagonal element equal to the number of neighbours of the area i, and the off-diagonal elements in each row equal to -1 if the corresponding areas are neighbours and 0 otherwise.
As the posterior distribution cannot be obtained analytically, full conditional distributions have been derived and samples from the posterior are obtained by Gibbs sampling.

3 Data description

The Italian territory covers an area of 301,340 km^2 dived into different classification levels according to the European Nomenclature of Territorial Units for Statistics (NUTS). For this analysis, the administrative province levels disaggregation (NUTS-3) is considered for a total amount of 107 units over the whole territory. At province level, data related to the production and collection of municipal solid waste (MSW) are retrieved from the waste national registry provided by the Italian Institute for Environmental Protection and Research, ISPRA (`http://www.catastorifiuti.isprambiente.it`). For 2019 and for each unit, the amount of MSW (in *tons*) and the amount of the differentiated fraction of MSW (DSW, in *tons*) are provided. This latter information refers to the organic and other materials (such as plastic, metal, glass, wood, paper and cardboard) fraction of generated MSW bound to recycling programmes and processed in composting facilities (including integrated anaerobic and aerobic treatment and anaerobic digestion treatment plants). In Italy as a whole for 2019, the production and collection of DSW represents the 61.35% of the total quantity of MSW with an amount generated per person of 306.29 Kg on average [3]. At province level for 2019, greater per capita productions of DSW ($\geq$ 480 Kg/inh) are observed in North-Central Italian units (*i.e.* Reggio Emilia, Rimini, Ferrara, Piacenza, Lucca). According to data availability and comparability at the Italian province level, potential determinants of per capita DSW generation are considered. In particular, factors related to demographic and socio-economic characteristics are evaluated: population (Pop), sex-ratio (SexR, as Male/Female), population density (PopDen, as pop/km^2), number of cities (Cit), average number of housing (nHous), unemployment rate (UnempR, as the % of not working people older than 15 years), graduates (Grad, as the % of people with higher educational degree), added value per capita (AddV in euros, as a proxy of the provincial Gross domestic product per capita). All indicators are provided by the the Italian Statistics Institute, ISTAT (available at `http://dati.istat.it/`). Besides, the number of plants where DSW are dumped for recycling, composting or treating by other methods is considered as potential factor affecting waste generation.

4 Preliminary results and further developments

The Hierarchical model in Eqs. 1 and 3 is fitted to evaluate the dependence relation between per capita production of DSW and the demographic and socio-economic covariates in Section 3. The joint posterior distribution of model parameters are obtained using 10,000 iterations discarding the first 1,000 using `R`'s `BayesSAE` package [1]. Model selection allows to determine relevant predictors as reported in Tab. 1. Estimated coefficients suggest positive effect of the population density and the added value covariates on the per capita DSW production; opposite signs are estimated for the number of housing and the unemployment rate implying higher production of differentiated waste fraction in wealthier provinces. The estimated mean spatial component in Fig. 1 shows the areal effect on the whole Italian territory highlighting the intrinsic correlation between similar provinces in terms of socio-economic characteristics: northern provinces are typically more productive and prosperous then the rest of the country. These preliminary results show that the spatial parametrization plays a significant role in modelling waste data. However, further discussion about the spatial component should be considered taking into account different diagnostic tools and alternative parameterizations of the spatial correlation.

Table 1: Posterior means of fixed effects and 95% Highest Posterior Density Intervals.

Parameter	mean	lower HPDI	upper HPDI
β_{PopDen}	0.144	0.058	0.217
β_{nHous}	−0.186	−0.275	−0.097
β_{UnempR}	−0.017	−0.033	−0.001
β_{AddV}	0.007	0.003	0.011

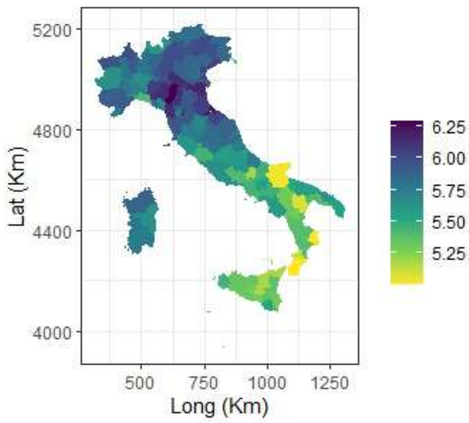

Fig. 1: Estimated mean effect of provinces on the DSW production per capita for 2019

References

1. Chengchun, S.: BayesSAE: Bayesian Analysis of Small Area Estimation. R package version 1.0-2. https://CRAN.R-project.org/package=BayesSAE, (2018)
2. Fay, R.E. and Herriot, R.A.: Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, **74**, 269–177, (1979)
3. ISPRA: Rapporti 331/2020. ISBN: 978-88-448-1030-6 (2020)
4. Keser, S., Duzgun, S., Aksoy, A.: Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. Waste Management **32**, 359–371 (2012)
5. Rao, J.N.K. and Molina, I.: Small Area Estimation, 2nd edn. Wiley, Hoboken, New Jersey (2015)
6. You, Y. and Zhou, Q.M.: Hierarchical Bayes small area estimation under a spatial model with application to health survey data. Survey Methdology, **37**, 1, 25–37, (2011)
7. Zhang, G., Lin, T., Chen, S. et al.: Spatial characteristics of municipal solid waste generation and its influential spatial factors on a city scale: a case study of Xiamen, China. J Mater Cycles Waste Manag, **17**, 399–409 (2015)

A spatial regression model for for predicting abundance of lichen functional groups

Un modello di regressione spaziale per la prevsione dell'abbondanza dei gruppi funzionali lichenici

Pasquale Valentini, Francesca Fortuna, Tonio Di Battista and Paolo Giordani

Abstract Lichen functional traits such as growth, photosynthetic and reproductive strategies, are considered indicators of changes in environmental conditions resulting especially from air pollution. However, due to the high variability of lichen flora, it is often difficult to differentiate the effects of atmospheric pollutants and those of other environmental variables. The aim of this paper is to evaluate the synergistic effect of atmospheric pollutants and environmental variables on lichen functional groups distribution through a Bayesian generalized spatial regression model.

Abstract *I tratti funzionali dei licheni, come la crescita, le strategie fotosintetiche e riproduttive, sono considerati importanti indicatori dei cambiamenti ambientali derivanti soprattutto dall'inquinamento atmosferico. Tuttavia, a causa dell'alta variabilita' della flora lichenica e' spesso difficile differenziare gli effetti degli inquinanti atmosferici da quelli di altre variabili ambientali. Questo articolo ha lo scopo di valutare l'effetto degli inquinanti atmosferici e delle variabili ambientali sulla distribuzione dei gruppi funzionali lichenici attraverso un modello di regressione bayesiano generalizzato.*

Key words: Bayesian spatial regression model, Lichen biodiversity, Air pollution, Biomonitoring techniques, Ecological functional traits

---

Pasquale Valentini
"G.d'Annunzio" University of Chieti-Pescara (Italy) e-mail: pasquale.valentini@unich.it

Francesca Fortuna
"Roma Tre" University, Rome (Italy) e-mail: francesca.fortuna@uniroma3.it

Tonio Di Battista
"G.d'Annunzio" University of Chieti-Pescara (Italy) e-mail: dibattis@unich.it

Paolo Giordani
"University of Genoa", Genoa (Italy) e-mail: giordani@dipteris.unige.it

1 Introduction

Lichens are particularly sensitive to environmental stresses due to their physiological characteristics and respond to phytotoxic gases at cellular, individual and community level [9]. However, different species react to pollutants in different ways, thus, it is often difficult to identify a direct and clear relationship between lichen biodiversity and pollution [2]. Moreover, the high variability of lichen diversity makes difficult to differentiate the effects of atmospheric pollutants and those of other environmental variables since lichen flora highly vary across geographic, climatic and ecological gradients [5].
For these reasons, approaches based on morpho-functional species traits (such as photosynthetic strategy, growth form or reproductive strategy) have been recently used to assess monitoring change in ecosystems [5, 8]. Functional traits allows to define functional groups, that is, groups of species that share some functional characteristics and, thus, react similarly to an environmental factor. However, the influence of environmental conditions on lichen functional traits is still poorly documented, hindering their use in environmental monitoring [5].
The aim of this paper is to examine the relationships between lichen functional groups and some covariates linked to environmental, habitat and air pollutant variables. At this purpose, we aim to determine the spatial pattern of lichens functional groups in response to a spatial gradient of pollution alteration. In this paper we introduce a lognormal and gamma mixed NB spatial regression model for counts [10]. In particular, the model proposed are presented in a hierarchical framework. This allows the inclusion of a "nugget" term in the spatial part of the model. The full model is estimated in a Bayesian setting and posterior inference is performed hierarchically via a Markov chain Monte Carlo scheme. The remainder of the paper proceeds as follows. In section 2 we briefly describe the data used in this study, while in section 3 we introduce the model. In section 4 we consider Bayesian inferential issues and, finally, in section 5 concludes the paper with a discussion.

2 The data

Epiphytic lichen biodiversity of Liguria region, in Northwestern Italy, has been considered. Data on lichen abundance has been collected following the standards suggested by [1]; the survey lasted from 2002 to 2003 and involved a total of 151 sampling sites of a 30 *m* radius plot (see Fig. 1) and 196 epiphytic lichen species i.e. lichens which live on trees bark [4].
Although the dataset does not refer to particularly up-to-date data, it has been selected as a model database to describe lichen colonisation processes under heterogeneous environmental conditions and included in The PREDICTS project (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems), which has collated representative database of comparable samples of biodiversity from multiple sites that differ in the nature or intensity of human impacts relating to land use [6].

UNIVERSITÀ DI PISA

Sant'Anna Scuola Universitaria Superiore Pisa

Consiglio Nazionale delle Ricerche

Università degli Studi di Cagliari

Book of Short Papers SIS 2021

SIS2021 PISA, 21-25 GIUGNO 2021

SIS Società Italiana di Statistica

Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Contents