

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 territoriali 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 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

$$y_i = X_i' \beta + v_i + e_i, i = 1, \dots, m, \quad (1)$$

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

$$\theta_i = X_i' \beta + v_i \quad (2)$$

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 $e_i \stackrel{\text{i.i.d.}}{\sim} N(0, \psi_i)$ and ψ_i a-priori known;
 Stage 2. $\theta_i = x_i^T \beta + v_i \quad i = 1, \dots, m$, with $v_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_v^2)$;
 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:

$$\theta_i = x_i^T \beta + b_i \quad (3)$$

where

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

and

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

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 i th 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 km² divided 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 (≥ 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/km²), 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

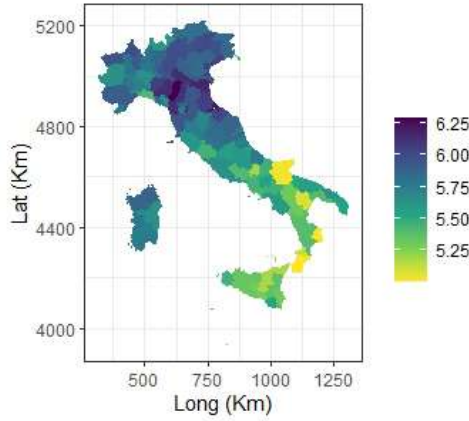


Fig. 1: Estimated mean effect of provinces on the DSW production per capita for 2019

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 Methodology*, **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 predicting abundance of lichen functional groups

Un modello di regressione spaziale per la previsione 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 variabilità 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].

Book of Short Papers

SIS 2021



Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo



Distribuzione Software | Formazione Professionale
Statistica | Economia | Finanza | Biostatistica | Epidemiologia
Sanità Pubblica | Scienze Sociali
www.tstat.it | www.tstattraining.eu

Copyright © 2021

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Contents

Preface	XIX
1 Plenary Sessions	1
1.1 Citizen data, and citizen science: a challenge for official statistics. <i>Monica Pratesi</i>	2
2 Specialized Sessions	8
2.1 A glimpse of new data and methods for analysing a rapidly changing population	9
2.1.1 The diffusion of new family patterns in Italy: An update. <i>Arnstein Aassve, Letizia Mencarini, Elena Pirani and Daniele Vignoli</i>	10
2.1.2 Causes of death patterns and life expectancy: looking for warning signals. <i>Stefano Mazzucco, Emanuele Aliverti, Daniele Durante and Stefano Campostrini</i>	16
2.2 Advances in ecological modelling	22
2.2.1 A Bayesian joint model for exploring zero-inflated bivariate marine litter data. <i>Sara Martino, Crescenza Calculli and Porzia Maiorano</i>	23
2.3 Advances in environmental statistics	29
2.3.1 Bayesian small area models for investigating spatial heterogeneity and factors affecting the amount of solid waste in Italy. <i>Crescenza Calculli and Serena Arima</i>	30
2.3.2 A spatial regression model for predicting abundance of lichen functional groups. <i>Pasquale Valentini, Francesca Fortuna, Tonio Di Battista and Paolo Giordani</i>	36
2.4 Advances in preference and ordinal data theoretical improvements and applications	42
2.4.1 Boosting for ranking data: an extension to item weighting. <i>Alessandro Albano, Mariangela Sciandra and Antonella Plaia</i>	43
2.4.2 An Extended Bradley-Terry Model For The Analysis Of Financial Data. <i>Alessio Baldassarre, Elise Dusseldorp and Mark De Rooij</i>	49

2.5	Business system innovation, competitiveness, productivity and internationalization	55
2.5.1	An analysis of the dynamics of the competitiveness for some European Countries. <i>Andrea Marletta, Mauro Mussini and Mariangela Zenga</i>	56
2.5.2	National innovation system and economic performance in EU. An analysis using composite indicators. <i>Alessandro Zeli</i>	62
2.6	Challenges for observational studies in modern biomedicine	68
2.6.1	Data integration: a Statistical view. <i>Pier Luigi Conti</i>	69
2.6.2	Exploring patients' profile from COVID-19 case series data: beyond standard statistical approaches. <i>Chiara Brombin, Federica Cugnata, Pietro E. Cippà, Alessandro Ceschi, Paolo Ferrari and Clelia di Serio</i>	75
2.6.3	On the statistics for some pivotal anti-COVID-19 vaccine trials. <i>Mauro Gasparini</i>	81
2.7	Data Science for Industry 4.0 (ENBIS)	87
2.7.1	Sample selection from a given dataset to validate machine learning models. <i>Bertrand Iooss</i>	88
2.7.2	Reliable data-drive modelling and optimisation of a batch reactor using bootstrap aggregated deep belief networks. <i>Changhao Zhu and Jie Zhang</i>	94
2.8	Integration of survey with alternative sources of data	100
2.8.1	A parametric empirical likelihood approach to data matching under nonignorable sampling and nonresponse. <i>Daniela Marella and Danny Pfeffermann</i>	101
2.8.2	Survey data integration for regression analysis using model calibration.	107
2.8.3	Latent Mixed Markov Models for the Production of Population Census Data on Employment. <i>Daniela Filippini, Ugo Guarnera and Roberta Varriale</i>	112
2.9	Media, social media and demographic behaviours	118
2.9.1	Monitoring the Numbers of European Migrants in the United Kingdom using Facebook Data. <i>Francesco Rampazzo, Jakub Bijak, Agnese Vitali, Ingmar Weber and Emilio Zagheni</i>	119
2.10	New developments in ensemble methods for classification	125
2.10.1	An alternative approach for nowcasting economic activity during COVID-19 times. <i>Alessandro Spelta and Paolo Pagnottoni</i>	126
2.10.2	Assessing the number of groups in consensus clustering by pivotal methods. <i>Roberta Pappadà, Francesco Pauli and Nicola Torelli</i>	132
2.10.3	Clustering of data recorded by Distributed Acoustic Sensors to identify vehicle passage and typology. <i>Antonio Balzanella and Stefania Nacchia</i>	138
2.11	New developments in latent variable models	144
2.11.1	A Hidden Markov Model for Variable Selection with Missing Values. <i>Fulvia Pennoni, Francesco Bartolucci, and Silvia Pandolfi</i>	145
2.11.2	Comparison between Different Likelihood Based Estimation Methods in Latent Variable Models for Categorical Data. <i>Silvia Bianconcini and Silvia Cagnone</i>	151
2.11.3	A Comparison of Estimation Methods for the Rasch Model. <i>Alexander Robitzsch</i>	157

2.12	New issues on multivariate and univariate quantile regression	163
2.12.1	Directional M-quantile regression for multivariate dependent outcomes. <i>Luca Merlo, Lea Petrella and Nikos Tzavidis</i>	164
2.13	Semi-parametric and non-parametric latent class analysis	170
2.13.1	Stepwise Estimation of Multilevel Latent Class Models. <i>Zsuzsa Bakk, Roberto di Mari, Jennifer Oser and Jouni Kuha</i>	171
2.13.2	Distance learning, stress and career-related anxiety during the Covid-19 pandemic: a students perspective analysis. <i>Alfonso Iodice D'Enza, Maria Iannario, Rosaria Romano</i>	177
2.13.3	A Tempered Expectation-Maximization Algorithm for Latent Class Model Estimation. <i>Luca Brusa, Francesco Bartolucci and Fulvia Pennoni</i>	183
2.14	Statistics for finance high frequency data, large dimension and networks	189
2.14.1	The Italian debt not-so-flash crash. <i>Maria Flora and Roberto Reno'</i>	190
3	Solicited Sessions	197
3.1	Advances in social indicators research and latent variables modelling in social sciences	198
3.1.1	A composite indicator to measure frailty using administrative healthcare data. <i>Margherita Silan, Rachele Brocco and Giovanna Boccuzzo</i>	199
3.1.2	Clusters of contracting authorities over time: an analysis of their behaviour based on procurement red flags. <i>Simone Del Sarto, Paolo Coppola and Matteo Troia</i>	205
3.1.3	An Application of Temporal Poset on Human Development Index Data. <i>Leonardo Salvatore Alaimo, Filomena Maggino and Emiliano Seri</i>	211
3.1.4	The SDGs System: a longitudinal analysis through PLS-PM. <i>Rosanna Cataldo, Maria Gabriella Grassia and Laura Antonucci</i>	217
3.2	Changes in the life course and social inequality	223
3.2.1	Heterogeneous Income Dynamics: Unemployment Consequences in Germany and the US. <i>Raffaele Grotti</i>	224
3.2.2	In-work poverty in Germany and in the US: The role of parity progression. <i>Emanuela Struffolino and Zachary Van Winkle Z.</i>	230
3.2.3	Parenthood, education and social stratification. An analysis of female occupational careers in Italy. <i>Gabriele Ballarino and Stefano Cantalini</i>	236
3.3	Composition in the Data Science Era	242
3.3.1	Can we Ignore the Compositional Nature of Compositional Data by using Deep Learning Approaches? <i>Matthias Templ</i>	243
3.3.2	Principal balances for three-way compositions. <i>Violetta Simonacci</i>	249
3.3.3	Robust Regression for Compositional Data and its Application in the Context of SDG. <i>Valentin Todorov and Fatemah Alqallaf</i>	255

3.4	Evaluation of undercoverage for censuses and administrative data	261
3.4.1	Spatially balanced indirect sampling to estimate the coverage of the agricultural census. <i>Federica Piersimoni, Francesco Pantalone and Roberto Benedetti</i>	262
3.4.2	Next Census in Israel: Strategy, Estimation and Evaluation. <i>Danny Pfeffermann</i>	268
3.4.3	Administrative data for population counts estimations in Italian Population Census. <i>Antonella Bernadini, Angela Chieppa, Nicola Cibella and Fabrizio Solari</i>	274
3.4.4	LFS non response indicators for population register overcoverage estimation. <i>Lorella Di Consiglio, Stefano Falorsi</i>	279
3.5	Excesses and rare events in complex systems	285
3.5.1	Space-time extreme rainfall simulation under a geostatistical approach. <i>Gianmarco Callegher, Carlo Gaetan, Noemie Le Carrer and Ilaria Prosdocimi</i>	286
3.6	Hierarchical forecasting and forecast combination	292
3.6.1	Density calibration with consistent scoring functions. <i>Roberto Casarin and Francesco Ravazzolo</i>	293
3.6.2	Forecasting combination of hierarchical time series: a novel method with an application to CoVid-19. <i>Livio Fenga</i>	298
3.7	Household surveys for policy analysis	304
3.7.1	Did the policy responses to COVID-19 protect Italian households' incomes? Evidence from survey and administrative data. <i>Maria Teresa Monteduro, Dalila De Rosa and Chiara Subrizi</i>	305
3.8	Learning analytics methods and applications	311
3.8.1	Open-Source Automated Test Assembly: the Challenges of Large-Sized Models. <i>Giada Spaccapanico Proietti</i>	312
3.8.2	How Much Tutoring Activities May Improve Academic Careers of At-Risk Students? An Evaluation Study. <i>Marta Cannistra, Tommaso Agasisti, Anna Maria Paganoni and Chiara Masci</i>	318
3.8.3	Composite-based Segmentation Trees to Model Learners' performance. <i>Cristina Davino and Giuseppe Lamberti</i>	324
3.8.4	Test-taking Effort in INVALSI Assessments. <i>Chiara Sacco</i>	330
3.9	Light methods for hard problems	336
3.9.1	Fast Divide-and-Conquer Strategies to Solve Spatial Big Data Problems. <i>Michele Peruzzi</i>	337
3.9.2	Application of hierarchical matrices in spatial statistics. <i>Anastasiia Gorshechnikova and Carlo Gaetan</i>	343
3.10	Management and statistics in search for a common ground (AIDEA)	349
3.10.1	Customer Segmentation: it's time to make a change. <i>Fabrizio Laurini, Beatrice Luceri and Sabrina Latusi</i>	350
3.10.2	Multivariate prediction models: Altman's ZScore and CNDCEC's sectoral indicators. <i>Alessandro Danovi, Alberto Falini and Massimo Postiglione</i>	356
3.10.3	Comparing Entrepreneurship and Perceived Quality of Life in the European Smart Cities: a "Posetic" Approach. <i>Lara Penco, Enrico Ivaldi and Andrea Ciacci</i>	362

3.10.4	The Relationship between Business Economics and Statistics: Taking Stock and Ways Forward. <i>Amedeo Pugliese</i>	368
3.11	Mathematical methods and tools for finance and insurance (AMASES)	373
3.11.1	On the valuation of the initiation option in a GLWB variable annuity. <i>Anna Rita Bacinello and Pietro Millosovich</i>	374
3.11.2	Modern design of life annuities in view of longevity and pandemics. <i>Annamaria Olivieri</i>	380
3.11.3	Risk Management from Finance to Production Planning: An Assembly-to-Order Case Study. <i>Paolo Brandimarte, Edoardo Fadda and Alberto Gennaro</i>	386
3.11.4	Some probability distortion functions in behavioral portfolio selection. <i>Diana Barro, Marco Corazza and Martina Nardonthors</i>	392
3.12	Multiple system estimation	398
3.12.1	Multiple Systems Estimation in the Presence of Censored Cells. <i>Ruth King, Oscar Rodriguez de Rivera Ortega and Rachel McCrea</i>	399
3.12.2	Bayesian population size estimation by repeated identifications of units. A semi-parametric mixture model approach. <i>Tiziana Tuoto, Davide Di Cecco and Andrea Tancredi</i>	405
3.13	Network sampling and estimation	411
3.13.1	Targeted random walk sampling. <i>Li-Chun Zhang</i>	412
3.13.2	Estimation of poverty measures in Respondent-driven sampling. <i>María del Mar Rueda, Ismael Sánchez-Borrego and Héctor Mullo</i>	418
3.13.3	Sampling Networked Data for Semi-Supervised Learning Algorithms. <i>Simone Di Zio, Lara Fontanella, Francesco Pantalone and Federica Piersimoni</i>	423
3.13.4	A sequential adaptive sampling scheme for rare populations with a network structure. <i>Emilia Rocco</i>	429
3.14	New perspectives on multidimensional child poverty	435
3.14.1	Estimating uncertainty for child poverty indicators: The Case of Mediterranean Countries. <i>Ilaria Benedetti, Federico Crescenzi and Riccardo De Santis</i>	436
3.14.2	Child poverty and government social spending in the European Union during the economic crisis. <i>Angeles Sánchez and María Navarro</i>	442
3.14.3	The Children's Worlds Study: New perspectives on children's deprivation research. <i>Caterina Giusti and Antoanneta Potsi</i>	448
3.14.4	The impact of different definition of "households with children" on deprivation measures: the case of Italy. <i>Laura Neri and Francesca Gagliardi</i>	454
3.15	Perspectives in social network analysis applications	460
3.15.1	A comparison of student mobility flows in Erasmus and Erasmus+ among countries. <i>Kristijan Breznik, Giancarlo Ragozini and Marialuisa Restaino</i>	461
3.15.2	Network-based approach for the analysis of LexisNexis news database. <i>Carla Galluccio and Alessandra Petrucci</i>	467
3.15.3	A multiplex network approach to study Italian Students' Mobility. <i>Ilaria Primerano, Francesco Santelli and Cristian Usala</i>	473
3.15.4	Ego-centered Support Networks:a Cross-national European Comparison. <i>Emanuela Furfaro, Elvira Pelle, Giulia Rivellini and Susanna Zaccarin</i>	479

3.16	Statistical analysis of energy data	485
3.16.1	Machine learning models for electricity price forecasting. <i>Silvia Golia, Luigi Grossi, Matteo Pelagatti</i>	486
3.16.2	The impact of hydroelectric storage in the Italian power market. <i>Filippo Beltrami</i>	492
3.16.3	Jumps and cojumps in electricity price forecasting. <i>Peru Muniaín, Aitor Ciarreta and Ainhoa Zarraga</i>	498
3.17	Statistical methods and models for the analysis of sports data	507
3.17.1	Football analytics: a Higher-Order PLS-SEM approach to evaluate players' performance. <i>Mattia Cefis and Maurizio Carpita</i>	508
3.17.2	Bayesian regularized regression of football tracking data through structured factor models. <i>Lorenzo Schiavon and Antonio Canale</i>	514
3.17.3	A dynamic matrix-variate model for clustering time series with multiple sources of variation. <i>Mattia Stival</i>	520
3.17.4	Evaluating football players' performances using on-the-ball data. <i>David Dandolo</i>	526
3.18	The social and demographic consequences of international migration in Western societies	532
3.18.1	Employment and job satisfaction of immigrants: the case of Campania (Italy). <i>Alessio Buonomo, Stefania Capecci, Francesca Di Iorio and Salvatore Strozza</i>	533
3.18.2	Social stratification of migrants in Italy: class reproduction and social mobility from origin to destination. <i>Giorgio Piccitto, Maurizio Avola and Nazareno Panichella</i>	539
3.19	Well-being, healthcare, integration measurements and indicators (SIEDS)	545
3.19.1	A Composite Index of Economic Well-being for the European Union Countries. <i>Andrea Cutillo, Matteo Mazziotta and Adriano Pareto</i>	546
3.19.2	Poverty orderings and TIP curves: an application to the Italian regions. <i>Francesco M. Chelli, Mariateresa Ciommi and Chiara Gagliarano</i>	552
4	Contributed Sessions	558
4.1	Advances in clinical trials	559
4.1.1	Quantitative depth-based [18F]FMCH-avid lesion profiling in prostate cancer treatment. <i>Lara Cavinato, Alessandra Ragni, Francesca Ieva, Martina Sollini, Francesco Bartoli and Paola A. Erba</i>	560
4.1.2	Modelling longitudinal latent toxicity profiles evolution in osteosarcoma patients. <i>Marta Spreafico, Francesca Ieva and Marta Fiocco</i>	566
4.1.3	Information borrowing in phase II basket trials: a comparison of different designs. <i>Marco Novelli</i>	572
4.1.4	Q-learning Estimation Techniques for Dynamic Treatment Regime. <i>Simone Bogni, Debora Slanzi and Matteo Borrotti</i>	578
4.1.5	Sample Size Computation for Competing Risks Survival Data in GS-Design. <i>Mohammad Anamul Haque and Giuliana Cortese</i>	584

4.2	Advances in neural networks	590
4.2.1	Linear models vs Neural Network: predicting Italian SMEs default. <i>Lisa Crosato, Caterina Liberati and Marco Repetto</i>	591
4.2.2	Network estimation via elastic net penalty for heavy-tailed data. <i>Davide Bernardini, Sandra Paterlini and Emanuele Taufer</i>	596
4.2.3	Neural Network for statistical process control of a multiple stream process with an application to HVAC systems in passenger rail vehicles. <i>Gianluca Sposito, Antonio Lepore, Biagio Palumbo and Giuseppe Giannini</i>	602
4.2.4	Forecasting air quality by using ANNs. <i>Annalina Sarra, Adelia Evangelista, Tonio Di Battista and Francesco Bucci</i>	608
4.3	Advances in statistical methods	614
4.3.1	Robustness of Fractional Factorial Designs through Circuits. <i>Roberto Fontana and Fabio Rapallo</i>	615
4.3.2	Multi-objective optimal allocations for experimental studies with binary outcome. <i>Alessandro Baldi Antognini, Rosamaria Frieri, Marco Novelli and Maroussa Zagoraiou</i>	621
4.3.3	Analysis of three-way data: an extension of the STATIS method. <i>Laura Bocci and Donatella Vicari</i>	627
4.3.4	KL-optimum designs to discriminate models with different variance function. <i>Alessandro Lanteri, Samantha Leorato and Chiara Tommasi</i>	633
4.3.5	Riemannian optimization on the space of covariance matrices. <i>Jacopo Schiavon, Mauro Bernardi and Antonio Canale</i>	639
4.4	Advances in statistical methods and inference	645
4.4.1	Estimation of Dirichlet Distribution Parameters with Modified Score Functions. <i>Vincenzo Gioia and Euloge Clovis Kenne Pagui</i>	646
4.4.2	Confidence distributions for predictive tail probabilities. <i>Giovanni Fonseca, Federica Giummolè and Paolo Vidoni</i>	652
4.4.3	Impact of sample size on stochastic ordering tests: a simulation study. <i>Rosa Arboretti, Riccardo Ceccato, Luca Pegoraro and Luigi Salmaso</i>	658
4.4.4	On testing the significance of a mode. <i>Federico Ferraccioli and Giovanna Menardi</i>	664
4.4.5	Hommel BH: an adaptive Benjamini-Hochberg procedure using Hommel's estimator for the number of true hypotheses. <i>Chiara G. Magnani and Aldo Solari</i>	670
4.5	Advances in statistical models	676
4.5.1	Specification Curve Analysis: Visualising the risk of model misspecification in COVID-19 data. <i>Venera Tomaselli, Giulio Giacomo Cantone and Vincenzo Miracula</i>	677
4.5.2	Semiparametric Variational Inference for Bayesian Quantile Regression. <i>Cristian Castiglione and Mauro Bernardi</i>	683
4.5.3	Searching for a source of difference in undirected graphical models for count data – an empirical study. <i>Federico Agostinis, Monica Chiogna, Vera Djordjilovic, Luna Pianesi and Chiara Romualdi</i>	689
4.5.4	Snipped robust inference in mixed linear models. <i>Antonio Lucadamo, Luca Greco, Pietro Amenta and Anna Crisci</i>	695

4.6	Advances in time series	701
4.6.1	A spatio-temporal model for events on road networks: an application to ambulance interventions in Milan. <i>Andrea Gilardi and Riccardo Borgoni and Jorge Mateu</i>	702
4.6.2	Forecasting electricity demand of individual customers via additive stacking. <i>Christian Capezza, Biagio Palumbo, Yannig Goude, Simon N. Wood and Matteo Fasiolo</i>	708
4.6.3	Hierarchical Forecast Reconciliation on Italian Covid-19 data. <i>Andrea Marconcchia, Serena Arima and Pierpaolo Brutti</i>	714
4.6.4	Link between Threshold ARMA and tdARMA models. <i>Guy M��lard and Marcella Niglio</i>	720
4.7	Bayesian nonparametrics	726
4.7.1	Bayesian nonparametric prediction: from species to features. <i>Lorenzo Masoero, Federico Camerlenghi, Stefano Favaro and Tamara Broderick</i>	727
4.7.2	A framework for filtering in hidden Markov models with normalized random measures. <i>Filippo Ascolani, Antonio Lijoi, Igor Pr��nster and Matteo Ruggiero</i>	733
4.7.3	On the convex combination of a Dirichlet process with a diffuse probability measure. <i>Federico Camerlenghi, Riccardo Corradin and Andrea Ongaro</i>	739
4.7.4	Detection of neural activity in calcium imaging data via Bayesian mixture models. <i>Laura D'Angelo, Antonio Canale, Zhaoxia Yu and Michele Guindani</i>	745
4.8	Clustering for complex data	751
4.8.1	Clustering categorical data via Hamming distance. <i>Edoardo Filippi-Mazzola, Raffaele Argiento and Lucia Paci</i>	752
4.8.2	Penalized model-based clustering for three-way data structures. <i>Andrea Cappelozzo, Alessandro Casa, and Michael Fop</i>	758
4.8.3	Does Milan have a smart mobility? A clustering analysis approach. <i>Nicola Cornali, Matteo Seminati, Paolo Maranzano and Paola M. Chiodini</i>	764
4.8.4	A Fuzzy clustering approach for textual data. <i>Irene Cozzolino, Maria Brigida Ferraro and Peter Winker</i>	770
4.8.5	Valid Double-Dipping via Permutation-Based Closed Testing. <i>Anna Vesely, Livio Finos, Jelle J. Goeman and Angela Andreella</i>	776
4.9	Data science for complex data	782
4.9.1	Text mining on large corpora using Taltac4: An explorative analysis of the USPTO patents database. <i>Pasquale Pavone, Arianna Martinelli and Federico Tamagni</i>	783
4.9.2	Emotion pattern detection on facial videos using functional statistics. <i>Rongjiao Ji, Alessandra Micheletti, Natasa Krklec Jerinkic and Zoranka Desnica</i>	789
4.9.3	The spread of contagion on Twitter: identification of communities analysing data from the first wave of the COVID-19 epidemic. <i>Gianni Andreozzi, Salvatore Pirri, Giuseppe Turchetti and Valentina Lorenzoni</i>	795
4.9.4	Composition-on-Function Regression Model for the Remote Analysis of Near-Earth Asteroids. <i>Mara S. Bernardi, Matteo Fontana, Alessandra Menafoglio, Alessandro Pisello, Massimiliano Porreca, Diego Perugini and Simone Vantini</i>	801
4.9.5	Determinants of football coach dismissal in Italian League Serie A. <i>Francesco Porro, Marialuise Restaino, Juan Eloy Ruiz-Castro and Mariangela Zenga</i>	805
4.10	Data science for unstructured data	810
4.10.1	Identification and modeling of stop activities at the destination from GPS tracking data. <i>Nicoletta D'Angelo, Giada Adelfio, Antonino Abbruzzo and Mauro Ferrante</i>	811

4.10.2	A generalization of derangement. <i>Maurizio Maravalle and Ciro Marziliano</i>	817
4.10.3	Analysis of clickstream data with mixture hidden markov models. <i>Furio Urso, Antonino Abbruzzo and Maria Francesca Cracolici</i>	823
4.10.4	Using Google Scholar to measure the credibility of preprints in the COVID-19 Open Research Dataset (CORD-19). <i>Manlio Migliorati, Maurizio Carpita, Eugenio Brentari</i>	829
4.10.5	Mobile phone use while driving: a Structural Equation Model to analyze the Behavior behind the wheel. <i>Carlo Cavicchia and Pasquale Samacchiaro</i>	835
4.11	Demographic analysis	841
4.11.1	Life expectancy in the districts of Taranto. <i>Stefano Cervellera, Carlo Cusatelli and Massimiliano Giacalone</i>	842
4.11.2	Family size and Human Capital in Italy: a micro-territorial analysis. <i>Gabriele Rui, Marco Breschi and Alessio Fornasin</i>	848
4.11.3	Estimate age-specific fertility rates from summary demographic measures. An Indirect Model Levering on Deep Neural Network. <i>Andrea Nigri</i>	854
4.11.4	Patterns in the relation between causes of death and gross domestic product. <i>Andrea Nigri and Federico Crescenzi</i>	860
4.11.5	Locally sparse functional regression with an application to mortality data. <i>Mauro Bernardi, Antonio Canale, Marco Stefanucci</i>	866
4.12	Environmental statistics	871
4.12.1	A Distribution-Free Approach for Detecting Radioxenon Anomalous Concentrations. <i>Michele Scagliarini, Rosanna Gualdi, Giuseppe Ottaviano, Antonietta Rizzo and Franca Padoani</i>	872
4.12.2	Ecosud Car, a novel approach for the predictive control of the territory. <i>Giacomo Iula, Massimo Dimo, Saverio Gianluca Crisafulli, Marco Vito Calciano, Vito Santarcangelo and Massimiliano Giacalone</i>	878
4.12.3	Effect of ties on the empirical copula methods for weather forecasting. <i>Elisa Perrone, Fabrizio Durante and Irene Schicker</i>	884
4.12.4	Spatio-temporal regression with differential penalization for the reconstruction of partially observed signals. <i>Eleonora Amone and Laura M. Sangalli</i>	890
4.12.5	Sea Surface Temperature Effects on the Mediterranean Marine Ecosystem: a Semiparametric Model Approach. <i>Claudio Rubino, Giacomo Milisenda, Antonino Abbruzzo, Giada Adelfio, Mar Bosch-Belmar, Francesco Colloca, Manfredi Di Lorenzo and Vita Gancitano</i>	895
4.13	Functional data analysis	901
4.13.1	Remote Analysis of Chapas Stops in Maputo from GPS data: a Functional Data Analysis Approach. <i>Agostino Torti, Davide Ranieri and Simone Vantini</i>	902
4.13.2	A Conformal approach for functional data prediction. <i>Jacopo Diquigiovanni, Matteo Fontana and Simone Vantini</i>	907
4.13.3	Block testing in covariance and precision matrices for functional data analysis. <i>Marie Morvan, Alessia Pini, Madison Giacomci and Valerie Monbet</i>	911
4.13.4	Analysing contributions of ages and causes of death to gender gap in life expectancy using functional data analysis. <i>Alessandro Feraldi, Virginia Zarulli, Stefano Mazzucco and Cristina Giudici</i>	917
4.13.5	Supervised classification of ECG curves via a combined use of functional data analysis and random forest to identify patients affected by heart disease. <i>Fabrizio Maturo and Rosanna Verde</i>	923

4.14	Mixture models	929
4.14.1	Alternative parameterizations for regression models with constrained multivariate responses. <i>Roberto Ascarì, Agnese Maria Di Brisco, Sonia Migliorati and Andrea Ongaro</i>	930
4.14.2	Spatially dependent mixture models with a random number of components. <i>Matteo Gianella, Mario Beraha and Alessandra Guglielmi</i>	936
4.14.3	Finite mixtures of regression models for longitudinal data. <i>Marco Alfò and Roberto Rocci</i>	942
4.14.4	Mixtures of regressions for size estimation of heterogeneous populations. <i>Gianmarco Caruso</i>	948
4.14.5	Finite mixtures of regressions with random covariates using multivariate skewed distributions. <i>Salvatore D. Tomarchio, Michael P.B. Gallaughier, Antonio Punzo and Paul D. McNicholas</i>	954
4.15	New applications of regression models	960
4.15.1	The Shapley-Lorenz decomposition approach to mitigate cyber risks. <i>Paolo Giudici and Emanuela Raffinetti</i>	960
4.15.2	A spatially adaptive estimator for the function-on-function linear regression model with application to the Swedish Mortality dataset. <i>Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo and Simone Vantini</i>	967
4.15.3	POSetR: a new computationally efficient R package for partially ordered data. <i>Alberto Arcagni, Alessandro Avellone and Marco Fattore</i>	972
4.15.4	Multi Split Conformal Prediction. <i>Aldo Solari and Vera Djordjilović</i>	978
4.15.5	Changes in the consumption of fruits and vegetables among university students during master courses: an analysis of data automatically collected from cashier transactions. <i>Valentina Lorenzoni, Giuseppe Turchetti and Lucio Masserini</i>	984
4.16	New challenges in clustering and classification techniques	990
4.16.1	A Dynamic Stochastic Block Model with infinite communities. <i>Roberto Casarin and Ovielt Baltodano López</i>	991
4.16.2	Cross-Subject EEG Channel Selection for the Detection of Predisposition to Alcoholism. <i>Michela Carlotta Massi and Francesca Ieva</i>	997
4.16.3	Some Issues on the Parameter Selection in the Spectral Methods for Clustering. <i>Cinzia Di Nuzzo and Salvatore Ingrassia</i>	1003
4.16.4	The link-match tale: new microdata from unit level association. <i>Riccardo D'Alberto, Meri Raggi and Daniela Cocchi</i>	1009
4.17	New developments in Bayesian methods	1015
4.17.1	Spatio-temporal analysis of the Covid-19 spread in Italy by Bayesian hierarchical models. <i>Nicoletta D'Angelo, Giada Adelfio and Antonino Abbruzzo</i>	1016
4.17.2	Modelling of accumulation curves through Weibull survival functions. <i>Alessandro Zito, Tommaso Rigon and David B. Dunson</i>	1021
4.17.3	Model fitting and Bayesian inference via power expectation propagation. <i>Emanuele Degani, Luca Maestrini and Mauro Bernardi</i>	1026
4.17.4	Bayesian quantile estimation in deconvolution. <i>Catia Scricciolo</i>	1032
4.17.5	Bayesian inference for discretely observed non-homogeneous Markov processes. <i>Rosario Barone and Andrea Tancredi</i>	1038