



SDS
Statistica e
Data Science



Palermo, 11-12 April 2024

Proceedings of the Statistics and Data Science 2024 Conference

New perspectives on Statistics and Data Science

Edited by

Antonella Plaia – Leonardo Egidi
Antonino Abbruzzo

Proceedings of the SDS 2024 Conference
Palermo, 11-12 April 2024
Edited by: Antonella Plaia - Leonardo Egidi - Antonino
Abbruzzo

-

Palermo: Università degli Studi di Palermo.

ISBN Ebook (Pdf)
978-88-5509-645-4

Questo volume è rilasciato sotto licenza Creative Commons
Attribuzione - Non commerciale - Non opere derivate 4.0



© 2024 The Authors

Contents

1	Keynote Sessions	11
1.1	Classification with imbalanced data and the (eternal?) struggle between statistics and data science. <i>Nicola Torelli</i>	11
1.2	Deep residual networks and differential equations. <i>G�rard Biau</i>	13
2	Invited - Complex data: new methodologies and applications ..	15
2.1	Link selection in binary regression models with the Model Confidence Set. <i>Michele La Rocca and Marcella Niglio and Marialuisa Restaino</i>	15
2.2	A cluster-weighted model for COVID-19 hospital admissions. <i>Daniele Spinelli, Paolo Berta, Salvatore Ingrassia and Giorgio Vittadini</i>	23
2.3	Multi-class text classification of news data. <i>Maurizio Romano and Maria Paola Priola</i>	28
3	Invited - Data science and dataspace: challenges, results, and next steps	35
3.1	Data-Centric AI : A new Frontier emerging in Data Science. <i>Donato Malerba, Vincenzo Pasquadibisceglie, Vito Recchia and Annalisa Appice</i>	35
3.2	Data Spaces strategy to unleash agriculture data value: a concrete use case. <i>Nicola Masi, Delia Milazzo, Giulia Antonucci and Susanna Bonura</i>	42
3.3	Addressing Agricultural Data Management Challenges with the Enhanced TRUE Connector. <i>Sergio Comella, Delia Milazzo, Mattia Giuseppe Marzano, Giulia Antonucci, Susanna Bonura and Angelo Marguglio</i>	48
4	Solicited - Data Science for Official Statistics	55
4.1	Data science at Istat for urban green. <i>Fabrizio De Fausti, Marco Di Zio, Giuseppe Lancioni, Stefano Mugnoli, Alberto Sabbi and Francesco Sisti</i>	55

4.2	Twitter (X) as a Data Source for Official Statistics: Monitoring Italian Debate on Immigration through Text Analysis. <i>Elena Catanese, Gerarda Grippo, Francesco Ortame and Maria Clelia Romano</i>	62
5	Solicited - Sustainable Artificial Intelligence in Finance	69
5.1	Feature Dependence and Prediction Explanations in P2P Lending. <i>Paolo Pagnottoni and Thanh Thuy Do</i>	69
6	Solicited - Young SIS	77
6.1	Merging data and historical information via optimal power prior selection in Bayesian models. <i>Roberto Macrì Demartino, Leonardo Egidi, Nicola Torelli and Ioannis Ntzoufras</i>	77
6.2	Hierarchical Mixtures of Latent Trait Analyzers with concomitant variables. <i>Dalila Failli, Bruno Arpino, and Maria Francesca Marino</i>	84
6.3	A Simultaneous Spectral Clustering for Three-Way Data. <i>Cinzia Di Nuzzo and Salvatore Ingrassia</i>	90
7	Solicited - From Data Analysis to Data Science	97
7.1	Optimal Scaling: New Insights Into an Old Problem. <i>Gilbert Saporta</i>	97
8	Solicited - Statistical methods for textual data	101
8.1	PROCSIMA: Probability Distribution Clustering Using Similarity Matrix Analysis. <i>Marco Ortu</i>	101
8.2	Exploring Anti-Migrant Rhetoric on Italian Social Media. <i>Lara Fontanella, Annalina Sarra, Emiliano del Gobbo, Alex Cucco and Sara Fontanella</i>	108
8.3	Causal inference from texts: a random-forest approach. <i>Chiara Di Maria, Alessandro Albano, Mariangela Sciandra and Antonella Plaia</i>	114
9	Solicited - Data analysis methods for data in non-Euclidean spaces	121
9.1	Riemannian Statistics for Any Type of Data. <i>Oldemar Rodriguez Rojas</i>	121
9.2	PAM clustering algorithm for ATR-FTIR spectral data selection: an application to multiple sclerosis. <i>Francesca Condino, Maria Caterina Crocco and Rita Guzzi</i>	128
9.3	Random Survival Forest for Censored Functional Data. <i>Giuseppe Loffredo, Elvira Romano and Fabrizio Maturo</i>	134
9.4	Advancing credit card fraud detection with innovative class partitioning and feature selection technique. <i>Mohammed Sabri, Antonio Balzanella and Rosanna Verde</i>	140
10	Solicited - Functional Data Analysis in Action	147
10.1	Functional Linear Discriminant Analysis for Misaligned Surfaces. <i>Tomas Masak</i>	147
10.2	Leveraging weighted functional data analysis to estimate earthquake-induced ground motion. <i>Teresa Bortolotti, Riccardo Peli, Giovanni Lanzano, Sara Sgobba and Alessandra Menafoglio</i>	155

10.3	Functional autoregressive processes on a spherical domain for global aircraft-based atmospheric measurements. <i>Almond Stöcker and Alessia Caponera</i>	161
11	Solicited - Bayesian Inference for Graphical Models	169
11.1	Log-likelihood approximation in Stochastic EM for Multilevel Latent Class Models. <i>Silvia Columbu, Nicola Piras and Jeroen K. Vermunt</i>	169
11.2	MCMC Sampling in Bayesian Gaussian Structure Learning. <i>Antonino Abbruzzo, Nicola Argentino, Reza Mohammadi, Maria De Iorio, Willem van den Boom and Alexandros Beskos</i>	176
12	Contributed - Promoting Equity: Statistical Insights into Tourism, Sustainability and Digital Divide	183
12.1	Lesson Learnt in the Data Science Worldview: New Dimension of Digital Divide. <i>Rita Lima</i>	183
12.2	An overview of Tourism Statistical Literacy. <i>Yasir Jehan, Giuseppina Lo Mascolo and Stefano De Cantis</i>	192
12.3	Scalable bootstrap inference via averaged Robbins-Monro approximations. <i>Giuseppe Alfonzetti and Ruggero Bellio</i>	198
12.4	The impact of sustainability on Initial Coin Offering: advantages in trading. <i>Alessandro Bitetto and Paola Cerchiello</i>	204
13	Contributed - High dimensional and functional data	211
13.1	Analysis of Brain-Body Physiological Rhythm Using Functional Graphical Models. <i>Rita Fici, Luigi Augugliaro and Ernst C. Wit</i>	211
13.2	A comparison of scalable estimation methods for large-scale logistic regression models with crossed random effects. <i>Ruggero Bellio and Cristiano Varin</i>	218
13.3	Single-cell Sequencing Data: Critical Analysis and Definition of Statistical Models. <i>Antonino Gagliano, Gianluca Sottile, Nicolina Sciaraffa, Claudia Coronello and Luigi Augugliaro</i>	224
13.4	Investigating the association between high school outcomes and university enrollment choice: a machine learning approach. <i>Andrea Priulla, Alessandro Albano, Nicoletta D'Angelo and Massimo Attanasio</i>	230
14	Contributed - Statistical Analysis in economic and market dynamics	237
14.1	A comparison of multi-factor stochastic models for commodity prices C3. <i>Luca Vincenzo Ballestra, Christian Tezza and Paolo Foschi</i>	237
14.2	Nonparametric ranking estimation with application to the propensity for Circular Economy of Italian economic sectors. <i>Stefano Bonnini, Michela Borghesi and Massimiliano Giacalone</i>	246
14.3	Impact of the Russian invasion of Ukraine on coal markets: Evidence from an event-study approach. <i>Yana Kostiuk, Paola Cerchiello and Arianna Agosto</i>	252
14.4	Labour market and time series: a forecast approach for European countries from 1995 to 2022. <i>Paolo Mariani, Andrea Marletta and Piero Quatto</i>	258

15	Contributed - Innovations in cluster and latent class models . . .	263
15.1	Biclustering of discrete data by extended finite mixtures of latent trait models. <i>Dalila Failli, Maria Francesca Marino and Francesca Martella</i>	263
15.2	Seismic events classification through latent class regression models for point processes. <i>Giada Lo Galbo, Giada Adelfio and Marcello Chiodi</i>	270
15.3	Determining the optimal number of clusters through Symmetric Non-Negative Matrix Factorization. <i>Agostino Stavolo, Maria Gabriella Grassia, Marina Marino and Rocco Mazza</i>	276
16	Contributed - Modelling on spatial phenomena	283
16.1	Integrating computational and statistical algorithms in RT-GSCS for spatial survey administration. <i>Yuri Calleo, Simone Di Zio and Francesco Pilla</i>	283
16.2	Sensitivity mapping as a tool to support siting of offshore wind farms and increase citizens' acceptability. <i>Giovanna Cilluffo, Gianluca Sottile, Laura Ciriminna, Geraldina Signa, Agostino Tomasello and Salvatrice Vizzini</i>	290
16.3	Investigating hotel consumers' purchase intention on web analytics data through PLS-SEM. <i>Giuseppina Lo Mascolo, Chiara di Maria, Marcello Chiodi and Arabella Mocciano Li Destri</i>	296
16.4	Spatio-temporal analysis of lightning point process data in severe storms. <i>Nicoletta D'Angelo, Milind Sharma, Marco Tarantino and Giada Adelfio</i>	302
17	Contributed - Statistical machine learning for predictive modelling	309
17.1	Application of statistical techniques to predict the effective temperature of young stars. <i>Marco Tarantino, Loredana Prisinzano and Giada Adelfio</i>	309
17.2	Topological Attention for Denoising Astronomical Images. <i>Riccardo Cecaroni and Pierpaolo Brutti</i>	316
17.3	LSTM-based Battery Life Prediction in IoT Systems: a proof of concept. <i>Vanessa Verrina, Andrea Vennera and Annarita Renda</i>	322
17.4	Predictive modeling of drivers' brake reaction time through machine learning methods. <i>Alessandro Albano, Giuseppe Salvo and Salvatore Rusotto</i>	328
18	Contributed - Ordinal and preference data analysis .	335
18.1	OSILA (Order Statistics In Large Arrays): an original algorithm for an efficient attainment of the order statistics. <i>Andrea Cerasa</i>	335
18.2	The Mallows model with respondents' covariates for the analysis of preference rankings. <i>Marta Crispino, Lucia Modugno and Cristina Mollica</i>	343
18.3	Value-Based Predictors of Voting Intentions: An Empirical and Explainable approach. <i>Luca Pennella and Amin Gino Fabbrucci Barbagli</i>	349
18.4	A dynamic version of the Massey's rating system with an application in basketball. <i>Paolo Vidoni and Enrico Bozzo</i>	355
19	Contributed - Advances in text mining	361
19.1	Can Correspondence Analysis Challenge Transformers in Authorship Attribution Tasks?. <i>Andrea Sciandra and Arjuna Tuzzi</i>	361

- 19.2 A Fuzzy Topic Modeling approach to legal corpora. *Antonio Calcagni and Arjuna Tuzzi* 368
- 19.3 EmurStat: a digital tool for statistical analysis of emur flow. *Simone Paesano, Maria Gabriella Grassia, Marina Marino, Dario Sacco and Rocco Mazza* 374
- 19.4 Graph Neural Networks for clustering medical documents. *Vittorio Torri and Francesca Ieva* 380

Seismic events classification through latent class regression models for point processes

Giada Lo Galbo, Giada Adelfio, Marcello Chiodi

Abstract We are trying to identify sub-processes of seismic events from the point processes' point of view and according to the latent class regression approach. Each seismic event is classified as membership of one of the 4 identified sub-classes of seismic sequences, each defined by particular and well-defined characteristics. So far, seismic sub-sequences have been identified and described according to several declustering methods. In this application, we show how sub-processes can be identified starting from the definition of a spatio-temporal intensity function for point processes, assuming independence of the past.

Key words: Latent class, mixture model, spatio-temporal point process, clustering, earthquake, seismic sequence

1 Introduction

The study of seismic sequences has been approached to identify sub-processes characterizing the background and the induced components [1]. What has been studied so far concerns the characteristics of the magnitude distribution corresponding to the cited above components [11], or the covariates affecting the induced component [3]. What has not been taken into account is the importance of covariates related to the occurrence of earthquakes. This study aims to identify the components of seismic processes, through the application of a latent class regression model, by analyzing the spatio-temporal intensity dependence on covariates.

Giada Lo Galbo · Giada Adelfio · Marcello Chiodi
Department of Economics, Business and Statistics, University of Palermo e-mail: giada.logalbo@unipa.it, e-mail: giada.adelfio@unipa.it, e-mail: marcello.chiodi@unipa.it

2 Methodology

2.1 Spatio-temporal point process

A spatio-temporal inhomogeneous Poisson point process has parametric probability density function, $f_X(X; \boldsymbol{\theta})$, defined within a spatio-temporal bounded window, $|W \times T|$, with volume $W > 0$ and length $T > 0$ as in Eq. (1)[4]:

$$f_X(X; \boldsymbol{\theta}) = e^{\int_W \int_T [1 - \lambda_{\boldsymbol{\theta}}(\mathbf{u}, t)] dt d\mathbf{u}} \prod_{(\mathbf{u}, t) \in X} \lambda_{\boldsymbol{\theta}}(\mathbf{u}, t) \quad \{(\mathbf{u}, t)\} \in X \quad \boldsymbol{\theta} \in \Theta \quad (1)$$

where: $\boldsymbol{\theta}$ is the set of regression parameters; Θ is the parameter space and $\lambda_{\boldsymbol{\theta}}(\mathbf{u}, t)$ is the parametric first-order intensity function describing the point pattern X .

Assuming log-linear dependence from a set of covariates, $\mathbf{Z}(\mathbf{u}, t)$, the first-order intensity function of the $(\mathbf{u}, t)^{th} \in X$ point is specified as in Eq. (2) [10]:

$$\lambda_{\boldsymbol{\theta}}(\mathbf{u}, t) = \exp \{ \boldsymbol{\theta}' \mathbf{Z}(\mathbf{u}, t) + \varrho(\mathbf{u}, t) \} \quad (2)$$

with $\varrho(\mathbf{u}, t)$ a scalar offset. The parameters, $\boldsymbol{\theta}$, are estimated by maximizing the log-likelihood function, $\ell_X(\boldsymbol{\theta}; X)$. Berman and Turner[5] propose a device, which adds a set of dummy points, X_d , within the spatio-temporal window, $|W \times T|$, where there are no observed points; i.e., the latter can be approximated as in Eq. (3):

$$\ell_X(\boldsymbol{\theta}; X) = \sum_{(\mathbf{u}, t) \in X_D} \ell_X(\boldsymbol{\theta}; (\mathbf{u}, t)) \approx \sum_{(\mathbf{u}, t) \in X_D} w_{(\mathbf{u}, t)} \{ y_{(\mathbf{u}, t)} \log [\lambda_{\boldsymbol{\theta}}(\mathbf{u}, t)] - \lambda_{\boldsymbol{\theta}}(\mathbf{u}, t) \} \quad (3)$$

where: $\sum_{(\mathbf{u}, t) \in X_D} w_{(\mathbf{u}, t)} = |W \times T|$; $X_D = X \cup X_d$; $w_{(\mathbf{u}, t)}$ is a quadrature weight of the $(\mathbf{u}, t)^{th}$ point; $y_{(\mathbf{u}, t)} = \frac{z_{(\mathbf{u}, t)}}{w_{(\mathbf{u}, t)}}$, where $z_{(\mathbf{u}, t)} = 1$ if $(\mathbf{u}, t) \in X$, $z_{(\mathbf{u}, t)} = 0$ if $(\mathbf{u}, t) \in X_d$.

2.2 Finite mixtures of log-linear regression models

A finite mixture model of R components has conditional density function, $g_X(X | \boldsymbol{\psi})$, defined by a whole set of parameters $\boldsymbol{\psi} = \{ \boldsymbol{\theta}_r, \pi_r \}_{r=1}^R$, as expressed in Eq. (4)[7, 8]:

$$g_X(X | \boldsymbol{\psi}) = \sum_{r=1}^R \pi_r f_X(X; \boldsymbol{\theta}_r) \quad \text{with: } \sum_{r=1}^R \pi_r = 1; \quad \pi_r > 0 \quad (4)$$

where: $\{ \boldsymbol{\theta}_r, \pi_r \}$ are the regression and weight parameters describing the r^{th} cluster.

The posterior probability of j^{th} cluster membership, is defined as in Eq. (5):

$$\rho_{(\mathbf{u}, t)_j} = P(j | (\mathbf{u}, t), \boldsymbol{\psi}) = \frac{\pi_j f_X((\mathbf{u}, t); \boldsymbol{\theta}_j)}{\sum_{r=1}^R \pi_r f_X((\mathbf{u}, t); \boldsymbol{\theta}_r)} \quad \forall (\mathbf{u}, t) \in X \quad (5)$$

where: $\{f_X(X; \boldsymbol{\theta}_r)\}_{r=1}^R$ is the set of components belonging to the finite mixture.

The parameters estimate, $\boldsymbol{\Psi}$, is obtained by maximizing the ‘complete data log-likelihood’, $c\ell_X(\boldsymbol{\Psi}; X)$, which is expressed as in Eq. (6):

$$c\ell_X(\boldsymbol{\Psi}; X) = \sum_{(\mathbf{u}, t) \in X} \log \left[\sum_{r=1}^R \pi_r f_X((\mathbf{u}, t); \boldsymbol{\theta}_r) \right] \quad (6)$$

The maximization of the complete data log-likelihood function is carried out through the iterative Expectation-Maximization (EM) algorithm [6]. For the r^{th} cluster, at the $(i+1)^{\text{th}}$ iteration, given: $\hat{\boldsymbol{\Psi}}^{(i)} = \{\hat{\boldsymbol{\pi}}^{(i)}, \hat{\boldsymbol{\theta}}^{(i)}\}$, the Expectation and the Maximization steps consist of, respectively [9]:

- **E-step** $\mathbb{E} \left[\hat{\pi}_r^{(i+1)} \right] = \frac{1}{n(X)} \sum_{(\mathbf{u}, t) \in X} \hat{\rho}_{(\mathbf{u}, t), r}^{(i)} = \frac{1}{n(X)} \sum_{(\mathbf{u}, t) \in X} \hat{P} \left(r \mid (\mathbf{u}, t), \hat{\boldsymbol{\Psi}}^{(i)} \right)$
- **M-step** $\hat{\boldsymbol{\theta}}_r^{(i+1)} = \arg \max_{\boldsymbol{\theta}_r} \left\{ \sum_{(\mathbf{u}, t) \in X} \hat{\rho}_{(\mathbf{u}, t), r}^{(i)} \ell_X \left(\boldsymbol{\theta}_r^{(i)}; (\mathbf{u}, t) \right) \right\}$

3 Application to seismic data

We are interested in classifying events characterizing a spatio-temporal clustered point pattern by a probabilistic clustering approach [8]; i.e., according to an estimated probability of latent class membership [2]. The data are provided by the Istituto Nazionale di Geofisica e Vulcanologia (INGV), and belong to the Catalogo delle Localizzazioni ASSolute (CLASS). They refer to Italian earthquakes that occurred between 1980 and 2018, with information on: depth, $D_p(\mathbf{u}, t)$; root mean square error from P/S arrival time, $R(\mathbf{u}, t)$; hypocentral error on vertical, $E_v(\mathbf{u}, t)$, and horizontal, $E_h(\mathbf{u}, t)$, coordinates; gap azimuth, $G(\mathbf{u}, t)$; distance from the nearest station, $D(\mathbf{u}, t)$; distance between probabilistic and expected hypocenter, $D(\mathbf{u}, t)$; quality location, $Q(\mathbf{u}, t)$; radius of a sphere with PDF volume, $P(\mathbf{u}, t)$.

By using the intensity function in Eq. (7), we fit models with a number of latent classes from $R = 2$ to $R = 8$ and choose the model with $R = 4$ latent classes, according to the AIC and BIC criteria:

$$\lambda_{\boldsymbol{\theta}}(\mathbf{u}, t) = e^{\{\theta_0 + \theta_1 D_p(\mathbf{u}, t) + \theta_2 R(\mathbf{u}, t) + \theta_3 E_v(\mathbf{u}, t) + \theta_4 E_h(\mathbf{u}, t) + \dots \\ \dots + \theta_5 G(\mathbf{u}, t) + \theta_6 D(\mathbf{u}, t) + \theta_7 L(\mathbf{u}, t) + \theta_8 Q(\mathbf{u}, t) + \theta_9 P(\mathbf{u}, t)\}} \quad (7)$$

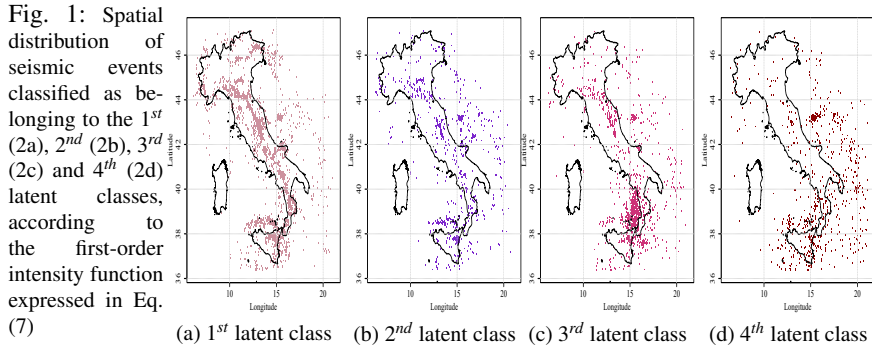
Tab. 1 shows the results of the latent class regression model fitting. As shown in Tab. 1, the covariates have the same positive (R, E_v, G, D) or negative (D_p, E_h, L, Q, P) effects, respectively, between the 1st and the 2nd latent classes, on the intensity of seismic events, although the different estimates’ values; some covariates have similar positive (D_p, G, D) or negative (Q, P) effects, on the intensity of seismic

events classified as belonging to the 3rd and the 4th latent classes. According to the regression parameters' estimates, the remaining covariates have null or opposite effects (and different magnitudes), among 3rd and 4th latent classes' intensities.

Table 1: Regression parameters estimates, $\hat{\theta}$, with standard errors, $\hat{\sigma}_{\hat{\theta}}$, and p -values, p , for the model with $R = 4$ latent classes and first-order intensity function defined as in Eq. (7)

r	θ_0	D_p θ_1	R θ_2	E_v θ_3	E_h θ_4	G θ_5	D θ_6	L θ_7	Q θ_8	P θ_9
$\hat{\theta}$	-14.677	-0.057	0.541	0.118	-0.003	0.008	0.011	-0.330	-3.086	-0.134
1 $\hat{\sigma}_{\hat{\theta}}$	0.094	0.004	0.038	0.014	0.008	0.001	0.002	0.021	0.240	0.023
p	< 0.001	< 0.001	< 0.001	< 0.001	0.718	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$\hat{\theta}$	-14.624	-0.044	0.153	0.204	-0.301	0.006	0.012	-0.199	-2.770	-0.170
2 $\hat{\sigma}_{\hat{\theta}}$	0.094	0.003	0.067	0.010	0.020	0.001	0.002	0.013	0.279	0.030
p	< 0.001	< 0.001	0.022	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$\hat{\theta}$	-14.788	0.004	0.089	0.062	0.056	0.011	0.005	-0.275	-4.657	-0.169
3 $\hat{\sigma}_{\hat{\theta}}$	0.096	0.001	0.034	0.006	0.007	0.001	0.001	0.018	0.255	0.021
p	< 0.001	< 0.001	0.009	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
$\hat{\theta}$	-14.946	0.006	-2.255	0.000	0.000	0.014	0.005	0.004	-4.205	-0.059
4 $\hat{\sigma}_{\hat{\theta}}$	0.115	0.000	0.281	0.001	0.001	0.001	0.001	0.001	0.256	0.007
p	< 0.001	< 0.001	< 0.001	0.847	0.977	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Figs. 1 and 2 show the spatial and the temporal distributions, respectively, of the 1st (2a, 3a), 2nd (2b, 3b), 3rd (2c, 3c) and 4th (2d, 3d) latent classes. As shown in Figs. 1 and 2, the 3rd and 4th latent classes are defined by seismic events loosely aggregated and nearer the tectonic zones around the Messina area. The 1st and 2nd latent classes identify seismic sequences which occurred in Central Italy and Northern Apennine chain.



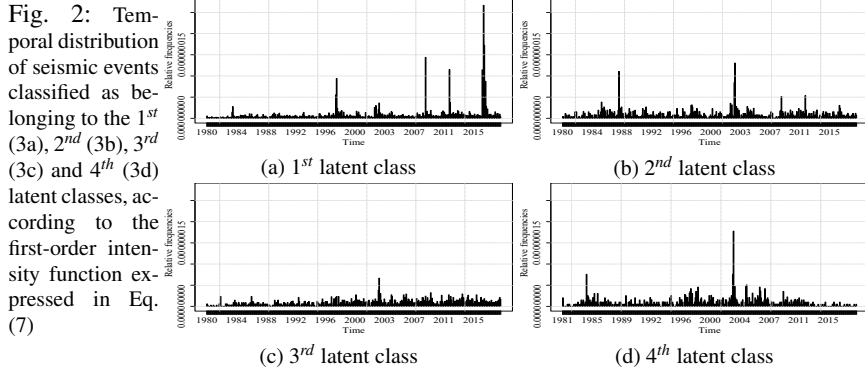


Fig. 3 and Tab. 2 show, respectively, the density functions of depth (D_p , 4d), magnitude (m , 4c), inter-time (D^T , 4a), inter-distance (D^S , 4b) and the weight parameters estimates, $\{\pi_r\}_{r=1}^R$, with the summary statistics of magnitude, m , depth, D_p , inter-time, D^T , and inter-distance, D^S , for each of the $R = 4$ latent classes.

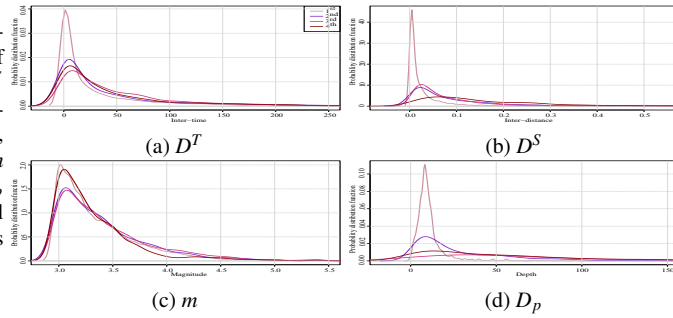
Table 2:

Weight parameters, $\{\pi_r\}_{r=1}^R$, and summary statistics of magnitude, m , depth, D_p , inter-time (hours), D^T , and inter-distance (km), D^S , for the model with $R = 4$ latent classes and first-order intensity function defined as in Eq. (7)

r	π_r	Quantile					Moment				
		Q_0	Q_1	Q_2	Q_3	Q_4	μ	σ	μ_3	β_2	
1	0.65	m	3.00	3.10	3.20	3.50	6.10	3.36	0.40	1.96	8.54
		D_p	-1.40	6.36	8.96	12.08	216.56	11.66	12.60	5.83	54.90
		D^T	0.00	0.62	8.18	38.15	1241.34	32.80	3749.80	4.94	52.49
		D^S	0.00	0.00	0.01	0.03	1.78	0.03	0.00	8.15	132.47
2	0.11	m	3.00	3.10	3.30	3.60	5.70	3.40	0.43	1.71	6.67
		D_p	-0.93	8.00	17.18	41.10	205.69	28.41	29.20	1.87	7.48
		D^T	0.00	2.76	17.15	58.85	1549.91	49.40	8070.56	6.53	87.12
		D^S	0.00	0.02	0.05	0.13	1.24	0.10	0.02	3.33	19.64
3	0.15	m	3.00	3.10	3.30	3.60	5.80	3.44	0.45	1.63	6.17
		D_p	-1.05	28.82	61.96	198.53	615.93	111.35	107.81	1.15	3.74
		D^T	0.00	8.08	29.81	69.92	690.53	52.50	4589.54	2.87	15.87
		D^S	0.00	0.02	0.05	0.11	1.89	0.10	0.02	4.77	38.36
4	0.09	m	3.00	3.10	3.20	3.50	5.40	3.32	0.37	2.11	8.72
		D_p	-0.69	13.98	40.94	82.42	558.78	64.81	75.59	2.48	11.47
		D^T	0.00	3.54	22.63	61.81	574.52	48.78	5057.82	2.71	12.80
		D^S	0.00	0.06	0.11	0.23	2.09	0.17	0.04	4.00	28.7

As shown in Fig. 3 and Tab. 2, the 1st and 2nd latent classes are in that order the most spatio-temporally aggregated and characterized by highest (lowest) maximum values of magnitude (depth). The 1st latent class is characterised by the majority of events belonging to the catalog. The 3rd and 4th latent classes of events are the least clustered and are characterized by events occurring at the highest depths. The latter groups of events correspond to deep earthquakes occurring around the Calabro-Ionian slab.

Fig. 3: Density function of inter-time, D^T (4a, hours), inter-distance, D^S (4b, km), magnitude, m (4c), and depth, D_p (4d), conditional to latent class membership



4 Conclusions

Despite the experimental purpose of the analysis, the application of a latent class regression model for point process, assuming independence of the past, allowed us to identify seismic sub-patterns. The results highlighted by the subsequent exploratory analysis on the events characterizing each latent class, are interesting. Further investigation of the identified sub-sequences, or accounting for different subsets of covariates, could lead to in-depth results concerning the seismic phenomenon.

References

- Adelfio, G., et al.: Earthquakes clustering based on maximum likelihood estimation of point process conditional intensity function. *Proceedings of the "The 4th International Workshop on Statistical Seismology" (StatSei4)*, **13**, (2006)
- Adelfio, G., Chiodi, M.: Alternated estimation in semi-parametric space-time branching-type point processes with application to seismic catalogs. *Stoch. Environ. Res. Risk Assess.*, **29**, pp. 443–450. Springer (2014)
- Adelfio, G., Chiodi, M.: Including covariates in a space-time point process with application to seismicity. *Stat. Methods. Appl.*, **30**, pp 947 – 971 (2021)
- Baddeley, A., Rubak, E., Turner, R.: *Spatial Point Patterns Methodology and Applications* with R. Chapman and Hall/CRC, New York (2015)
- Berman, M., Turner, T.R.: Approximating point process likelihoods with GLIM. *Appl. Stat.*, **41**:31 – 38, Wiley (1992)
- Dempster, A. P., Nan M. L., and Donald B. R.: Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Series B Stat. Methodol.* **39**:1, pp. 1 – 22 (1977)
- DeSarbo, W. S., and William L. C.: A maximum likelihood methodology for clusterwise linear regression. *J. Classif.*, **5**, pp. 249 – 282 (1988)
- Fraley, C., Raftery, A. E.: Model-Based Clustering, Discriminant Analysis, and Density Estimation. *J. Am. Stat. Assoc.*, **97**:458, pp. 611–631. Taylor & Francis (2002)
- Leisch, F.: *FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R*. *J. Stat. Softw.* **11** (2004)
- Lewis, P., A. W.: Recent results in the statistical analysis of univariate point processes. *N. P. S.* (1971)
- van Stiphout, T., Zhuang, J., Marsan, D.: Seismic declustering, Community Online Resource for Statistical Seismicity Analysis. (2012)