Statistics, Technology and Data Science for Economic
and Social Development

Book of short papers of the ASA Rome Conference
18 to 20 September 2024

“Measuring and Interpreting World Changes with Statistics,
Data Science and AI”

Book of Short Papers of the
Sapienza University of Rome Conference

Supplement to Volume 37/2

Guest Editors:
Leonardo S. Alaimo, Fabio Crescenzi,
Luigi Fabbris, Filomena Maggino, Maurizio Vichi

Statistica Applicata – ITALIAN JOURNAL OF APPLIED STATISTICS

A.S.A CONTACTS
Principal Contact
Francesco Palumbo (Editor in Chief)
editor@sa-ijas.org

Support Contact
Domenico Vistocco (Editorial Manager)
ijas@sa-ijas.org

JOURNAL WEBPAGE
https://www.sa-ijas.org/ojs/index.php/sa-ijas

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

SINERGIE – Italian Journal of Management

Pubblished: April 2025

Published by Firenze University Press
Powered by Firenze University Press

Firenze University Press
Università degli Studi di Firenze
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising:Cleup sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.
Rules for manuscript submission: https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions
Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA, 7 March 2023.

Summary

Statistica Applicata
ITALIAN JOURNAL OF APPLIED STATISTICS

Introduction 13

Determinants of cooperation among hotels in Venice: A network approach 15
Amati, V., Moretti, A., Finzi, R.M.

Modelling default risk in Italian SMEs: A spatial probit approach 21
Amoruso, E., Bruno, E., Castellano, R., Punzo, R.

The influence of UNESCO sites on tourism demand and supply: Insights from Italian regions .25
Antolini,F., Cesarini, S., Terragli, I.

“Inhumane”: Quantitative textual analysis of a qualitative survey on gender violence 31
among victims of international human trafficking
Arcaio, M., Barbaro, F., Parroco, A.M.

Keyword-enhanced summarization from scientific literature through an integrated approach .37
Aria, M., Cuccurullo, M., D’Aniello, L., Misuraca, M., Spano, M.

A composite indicator to measure attitude towards sustainable tourism by European citizens ..43
Bassi, F., Scioni, M.

Data quality assessment framework for composite indicator construction: An application to service delivery 49
Bassi, F., Muhanguzi, H., Scioni, M.

Exploring art market trends: A Delphi study on collectible design 55
Benedan, L., Codignola, F., Galeone, C., Mariani, P.

Resilience versus depression of Italian youth after the COVID-19 pandemic 61
Bollani, L., Cascella, C., Fabbris, L.

Trust in health science and organizations in Italy during the Covid-19 pandemic 67
Bollani, L., Fabbris. L.,

Artificial intelligence and future scenarios: Preliminary results from an application on family policies in north-eastern Italy 73
Bolzan, M., Calleo, Y., Di Zio, S., Marozzi, M., Scioni, M.

A non-parametric test for comparative evaluations of students' performances 79
Bonnini, S., Borghesi, M., Giacalone, M., Piscopo, G.

An empirical analysis of various methods for making inferences with non-probability Samples85
Braito, L., Rocco, E.

A statistical approach for the predictive analysis of insolvency in Italy 91
Brentari, E., Ermici, G., Zanetti, A.

An innovative point of view for statistical dissemination to prevent road accidents using the 'collision matrix'..97
Broccoli,M., Bruzzone, S.

A new algorithm for the correction and control of road accidents data based on string similarities...103
Bruzzone, S., Valentino, L., Moretti, D.,Tavanti, M., De Blasi, F.

Advancing spatial scenarios adopting generative pre-trained transformers and text-to-image models..109
Calleo, Y., Pilla, F., Di Zio, F.

Enhancing mobility and tourism through data analytics and generative AI 115
Cappellari, D.,Mantegar,G., Zamengo, B.

An innovative approach for the analysis of socio-economic phenomena at sub-municipal level: The Household Deprivation Study project...121
Carbonetti, G., Biasciucci, F., Cutillo, A., Mazziotta,M., Quondamstefano, V., Tronu, D.

Volatility decomposition of traffic flow time series with complex seasonality using GARCH models..127
Carpita, M., Metulini, R., Migliorati, M.

Economic well-being of households and attractiveness of territories: The case of Italian large municipalities...133
Casacci, S.

An application of the partial credit model to measure the resilience of Italian young adults in the aftermath of the COVID-19 pandemic 139
Cascella, C., Bollani, L., Fabbris, L.

Data science approaches to the study of multimorbidity patterns at older ages..........................145
Caselli, N., Dang, L., H., K., Rettaroli, L., Miglio, R.

Analysing fan engagement and sentiment on social media following Bologna FC's 2023/2024 Champions League qualification 151
Cavrini, G., Stavolo, A.

Text mining of tourist reviews: Sentiment analysis and trends in the Apulian tourism.............157
Centorame, D., Pino, G., Sarra, A., Nissi, E.

An exploratory analysis of the comments from the Istat's household budget survey questionnaires for designing an App 163
Cezza, E., D'Ambrosio, G., De Gaetano, L., Raimondi, E.

Combining expert knowledge and data-driven classification to support prediction and validation of census population counts....................169
Chieppa, A., Bernardini, A., Cibella, N., Solari, N.

The "group approach" versus the "inverted cascade approach" for calculating the super-dividend test of holding entities....................175
D'Achille, A., Boscaino, F.,

Migration as a diffusion process: The case of Bologna municipality....................181
D'Attoma, I., Cid, C., Tassinari, C.

Users beyond "boundaries": Recent experiences of promoting statistical literacy by Istat at the territorial level among non-regular data users....................187
Da Valle, S., Osti, S., Viviano, L.C.M.

Conceptualizing and measuring educational poverty through family, school, and environmental opportunities....................193
Davino, C., De Falco, A., Fabbricatore, R., Gherghi, M., Romano, R.

The Italian NEETs after the COVID-19: What should we expect for the future? Different approaches to forecasts....................199
De Luca, G., Rocca, A., Mazzocchi, P., Quintano C.

AI and geostatistics: Innovative solutions for reducing and storing CO2....................205
Digrandi, A., M., Cimmino, P.

Leveraging Industry 5.0 for official business statistics innovation....................211
Distefano, D., Bosso, P., Di Paolo, G., G., Papa, P.

Future orientation and psychological and social capital of Italian young adults....................217
D'Uggento, A.M., Fabbris, L.

Travel and vacations of Italian senior tourists....................223
D'Uggento, A.M., d'Ovidio, F., D, Marin, C., Pepe, I., Toma, E.

It's how you use the items that counts: An intelligent procedure for item selection in Item response theory....................229
Epifania,O., M., Anselmi,P., Robusto, E.

School socioeconomic status and students' achievements: Exploring territorial heterogeneity through a multilevel regression discontinuity approach....................235
Fabbricatore, R., Sannino, P., Davino, C.

Capacity to aspire in high school students: Theoretical definition and measurement....................241
Fabbricatore, R., Romano, R., Davino, C.

Improving online misogyny detection: The role of human annotation and explanation guided learning....................247
Fontanella, L., Cucco, A., del Gobbo, E., Ignazzi, E., Fontanella, S.

The evolution of Explainable Artificial Intelligence (XAI): A preliminary systematic literature view....................253
Gnasso, A., Aria, M.

Innovative ER solutions: Using data to drive operational success 259
Grassia, M.G., Sacco, D., Paesano, D., Dagostino, F., Annachiarico, A., Di Palma, V.

Responsible AI/ML implementation in official statistics: A framework for decision-making.....265
Grippo,· G., Righi, A.

Copula functions and dynamic hydrological data: The case of Scotland 269
Guerini, S.S., Spezia, L.

AI-generated sentiment in state-space models for Toyota Camry US volume sales forecasting 275
Iezzi, D., F., Monte, R., Pasquini, D.

Modeling virality in social media...281
Iezzi, D., F., Monte,R., Pasquini, D.

Organizational flexibility measures: Remote working at the Italian National Institute of Statistics 287
Limiti, C., Sola G.

Improving information on labour demand for decision-making by integrating traditional surveys, administrative sources, and web-based data 293
Lucarelli, A., Righi, A., Rizzi· R.

Early detection of university dropouts: An experiment using video-lectures access data 299
Martini, M.C.

Statistical indicators for the analysis of social sustainability and human well-being..................305
Massari, A., Perchinunno, P., L'Abbate, S., Leogrande, D., Mongelli, D.

Health and discrimination among refugees in Germany 311
Mendola, D., Busetta, A.

Estimating overdispersion in non-transformed scores of ANVUR's department ranking procedure: A Beto-multinomial model...317
Montanari,G., E., Doretti, M.

Design of a test for assessing teamwork and collaborative skills with competence-based knowledge space theory...323
Morleo, F., Vitanza, A., Anselmi, P.

Efficiency of Italian courts: Awareness from spatial stochastic frontier analysis 329
Nissi, E., Sarra, A., Evangelista, A., Di Battista, T.

Campanian healthcare facilities under the microscope: Reviews, polarity, and georeferencing335
Paesano, S., Grassia, M.G., Marino, M., Mazza, R., Sacco, D., Stavolo, A.

Markov switching autoregressive models and Bayesian hydrology 341
Paroli, R., Spezia, L.

Collecting ADS-B data for timely production of air traffic statistics 347
Pietropaoli, S., Amarone, M.

A statistical approach to measuring the impact of medication use on the lived experience of patients with inflammatory bowel disease 353
Piscitelli, A., D'Uggento, A., M., Equatore, E., Principi, M.B.

Artificial intelligence in support of accessibility: The Brain-friend device 359
Santarcangelo, V., Crisafulli, S.G., Giacalone, M., Piscopo, G., Massa, E., Vitullo, S., Scintu, D., Romano, A., Radosti, G.

Combining textual analysis and IRT to interpret the students' voice 365
Sarra, A., Fontanella, L., Evangelista, A., Di Battista, T.

Exploring teachers' professional identity: A case study utilizing structural equation modeling and text mining in a multimodal approach 371
Sarra,A., Pentucci, M.,Fontanella. L.

Selecting sample households with high quality addresses for population census: Problems for field survey network, but bias in the estimates?...377
Scavalli, E.

Statistical analysis of smartphone mobility data for air quality assessment 383
Shaukat, M., H., Finazzi, F.

The combination of expert knowledge and Bayesian networks to identify the relevant factors in diabetic kidney disease (DKD)...389
Slanzi, D., Poli, I., Jones, R., D.,Mayer, G.

Understanding consumer behavior in the plant-based meat market: Insights from Bayesian network analysis...395
Slanzi, D., Lanzini, P., Checchinato, F., Colapinto, C.

Exploring childfree and childless on social media: A text analysis approach 401
Stavolo, A., Cavrini, G., Cisotto, E., Marino, M.

Mass tourism and local housing: An exploratory textual analysis 407
Stavolo, A., Cataldo, R., Grassia, M., G., Marino, M.

Multiverse analysis: Heuristic for multi-model inference or theory of the model error?...........413
Tomaselli, V., Cantone, G., G.

From crisis to crisis: Mapping material deprivation in the EU through the 2000s 419
Traversa, S., Ivaldi, E.

The effect of user web journey on browsing behaviour and buying attitude: The case of Sunweb.nl...425
Urso, F., Argentino, N., Abbruzzo, A., Mohammadi, R., Pak, K., Cracolici, M., F.

Geography of business alliances and spatial network complexity: The backbone of formal collaboration agreements 429
Vurro, A., E., Bumbea, A., Giuffrida, A., Mazzitelli, A., Espa, G.

Spatial analysis of climate change awareness across countries **435**
Zammarchi, G., Maranzano, P.

Future scenarios: A financial network analysis of stock prices in emerging metaverse market..**441**
Zavarrone, E., Forciniti, A.

doi.org/10.26398/asaproc.00134

Statistical analysis of smartphone mobility data for air quality assessment

Muhammad Haroon Shaukat [a], Francesco Finazzi [b]
[a] Department of Engineering and Applied Sciences, University of Bergamo, Italy.
[b] Department of Economic Sciences, University of Bergamo, Italy.

1. Introduction

Population growth and urbanization have led to widespread pollution from industry, waste, heating, and traffic, causing cardiovascular and respiratory diseases (Yurtseven et al., 2018). To address this, personal exposure must be assessed by combining environmental data with individual mobility tracking. High-resolution location data improves exposure accuracy, linking pollution to health impacts more effectively (Finazzi and Paci, 2019).

Air quality is particularly crucial in cities like Istanbul, which has the highest population and busiest traffic in Turkey. Over the last five decades, the population of Istanbul has expanded sixfold, transforming it into a major metropolis both regionally and globally. Ministry of Environment and Urban Development of Turkey highlighted that gas emissions in the city exceed recommended levels. The primary contributors to this pollution are vehicular traffic and industrial activities. Additionally, Istanbul's location, which connects continents, results in heavy maritime traffic, further exacerbating emission levels (Bucak et al., 2021). Interestingly, restrictions on human mobility during the COVID-19 pandemic not only helped reduce the virus's spread but also improved air quality by curbing industrial production, transportation, and traffic (Zhu et al., 2020). This study examines the link between human mobility and air quality, focusing on the mediating role of individual movement patterns.

2. Data

This study analyses particulate matter with a diameter of 10 micrometers PM_{10} data obtained from an open-access platform[1], and the dataset covers measurements from thirteen air quality monitoring stations across Istanbul, including locations such as Aksaray, Besiktas, Esenyurt, Kadikoy, Kartal, Esenler, Sariyer, Sultangazi, Sile, Silivri, Umraniye, Uskudar, and Basaksehir. The data spans from March 1, 2023, to April 30, 2023, consisting of daily observations, recorded in micrograms per cubic meter. Secondly, the ERA5-Land dataset, with a 9x9 km latitude-longitude grid, was used due to its accurate representation of historical climate conditions. This study uses four climate variables covering the area around Istanbul's air quality monitoring stations, aligned with the air quality data, obtained from the Copernicus Climate Change Service[2]. Additionally, the individual smartphone mobility dataset of 443,786 users is considered, covering the same time period as previously mentioned, and was obtained from the Earthquake Network Project (www.sismo.app). This project operates a global earthquake early warning system using networks of smartphones (Finazzi, 2016).

3. Methodology

3.1 Fixed rank kriging

[1] https://aqicn.org/historical/#city:turkey/istanbul
[2] https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview

Fixed rank kriging (FRK) is a scalable spatio-temporal modeling and prediction framework that deals with large datasets and can easily handle data with varying spatial supports. FRK relies on a spatial random effects (SRE) model, where a spatially correlated mean-zero random process is decomposed into a linear combination of spatial basis functions with random coefficients, along with an additional term to capture the fine-scale variation of the random process. Let $\{Y(s): s \in D\}$ denote the spatial process, where s indexes represent the location within the spatial domain D. According to Zammit-Mangion and Cressie (2017), the classical spatial statistical model is as follows,

$$Y(s) = t(s)^T \alpha + \upsilon(s) + \xi(s), \quad s \in D \tag{1}$$

where s belongs to the spatial domain, $t(s)$ represents the spatially referenced covariates, α is associated regression coefficients with covariates, $\upsilon(s)$ indicates the small-scale, spatially correlated random effect, and $\xi(s)$ is a fine-scale almost spatially uncorrelated random effect. The latent process' structure can be described in terms of a linear combination of a fixed number of spatial basis functions which constitutes the SRE model as follows:

$$\lambda(s) = \sum_{l=1}^{r} \phi_l(s)\eta_l + \xi(s), \quad s \in D \tag{2}$$

where $\eta = (\eta_1,...,\eta_r)$ is an r-variate random vector, and $\phi(.) = (\phi_1(.),...,\phi_r(.))$ indicating the r-dimensional vector of pre-specified spatial basis functions (SBF). Further, the comprehensive information on spatial domain discretization into basic areal units (BAUs), parameter estimation, and the observational equation can be found in Zammit-Mangion and Cressie, (2017).

3.2 Performance measurement

A common approach for validating the accuracy of spatial-temporal models is to use the leave one out cross validation (LOOCV) approach. This method involves sequentially omitting data from one location (typically a station) at a time and estimating the value at that omitted location using the remaining data. According to Robinson and Metternicht (2006), the root mean square error (RMSE) is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}[Y(s_i,t_i) - \hat{Y}(s_i,t_i)]^2}{M}}, \tag{3}$$

where $(M; i = 1,...,M)$ indicating the total number of observations, $Y(s_i,t_i)$ are the actual space-time points, and $\hat{Y}(s_i,t_i)$ are the predicted space-time points.

4. Results and discussion

In the FRK setting, the first step is to create the BAUs N which can be either equally sized grids or hexagons. BAUs partition the spatial domain into discrete regions, allowing the fine-scale variation of latent processes to be modelled effectively. This approach allows to one use internal covariates or external. Therefore, the external covariates are assumed to be associated with each BAU, and the fine-scale variability can be incorporated in the form of weighted considering the variability of spatial latent processes. As far as the weights need to be equal to the number of BAUs.

The dimensions of the BAUs are considered to be 7x7 km, and therefore 124 BAUs are created per day. Consequently, over 61 days, a total of 7564 BAUs were obtained. The next step involves integrating ERA5-Land variables, available at a 9x9 km resolution, with the BAUs at a 7x7 km resolution, and this integration was achieved using a K-Nearest Neighbors (KNN) interpolation approach combined with inverse distance weighting (IDW). Additionally, BAUs are temporally replicated for each unique day to enable temporal interpolation. With 124 BAUs per day over 61 days, this results in a total of 7564 BAUs. For each day, ERA5-Land variables are interpolated for each BAU using IDW, where closer neighbors exert more influence on the interpolated values. This method ensures that both spatial and temporal correlations are effectively captured in the interpolation process. To include individual human mobility patterns as a covariate in the FRK model, a threshold of 3.5 km was applied. Since the BAUs have a uniform resolution of 7x7 km, any time a smartphone user passed within a 3.5 km radius of a BAU, it was considered that the user was present or had passed through that area on a specific date. This presence was marked with a "1" (indicating presence), while all other locations where the user neither passed through nor stayed within the 3.5 km radius of BAU were marked with a "0" (indicating absence). If a smartphone user passed through or stayed within the specified 3.5 km radius at the same location multiple times in a day, their presence was recorded cumulatively for that location and date. This process was repeated for each date to track the presence and absence of smartphone users across the spatial area and this covariate is reflected in the study as the presence of smartphone users.

It is important to describe that this study uses the three different types of basis functions such as bisquare, exponential, and Matern32, and for different combinations of basis function models, the centers of the temporal basis functions (TBFs) are set regularly using a sufficiently large-scale parameter ($\sigma = 2$) days and resulting 30 TBFs. For spatial basis functions (SBFs), two resolutions were considered with fixed scale parameters: a radius of 25 km for the first resolution and a radius of 15 km for the second resolution. SBFs are irregularly distributed across the entire geographical area of Istanbul, Turkey, covering the entire spatial domain of interest. As a result, 37 SBFs were created, with 4 corresponding to the first resolution and 33 to the second resolution. The use of two levels of spatial resolution in the basis functions can enhance the capability to model spatial correlations. Generally, a smaller resolution can capture local correlations, while a larger resolution can model more general correlations.

Afterward, both basis functions were combined for the spatial-temporal setting. This was achieved using the tensor product, resulting in 1110 combined spatial-temporal basis functions r obtained. Now, there are 5551 observations represented by footprints m with each footprint associated with a specific measurement error σ_{ε}^2. Nevertheless, the uncertainty for each measurement is assumed to be uniform across all footprints, set at a value of one, ensuring consistency throughout the spatial domain. Further, the expected maximization (EM) algorithm is employed with 30 iterations with a tolerance of 0.1 to estimate parameters. These settings are generally applied to fit all seven FRK models, each involving different combinations of basis functions. To ensure the validity of all seven models, the study employs the LOOCV criteria and then root mean square error (RMSE) was computed between the actual and predicted observations for each station and also computed average RMSE for all models. Further detail is described in Table 1.

From Table 1, it can be observed that the average RMSE shows small differences across all models, but the FRK model with Bisquare spatial and temporal basis function demonstrated the highest accuracy as compared to other combinations of models. Ultimately, this model was chosen to be applied across the entire dataset for predicting PM_{10} concentrations at unobserved locations in Istanbul, Turkey. The remaining steps of the model development followed the previously described procedure. The estimated intercept value is 30.51 $\mu g / m^3$. Since the covariates are externally linked to the BAUs, the FRK model does not provide parameter estimates for the covariates.

Table 1: Comparison of FRK models by LOOCV w.r.t to spatial-temporal basis function

Model	M1	M2	M3	M4	M5	M6
Spatial and Temporal Basis Function	Bisquare Spatial and Bisquare Temporal Basis Functions	Bisquare Spatial and Exponential Temporal Basis Function	Bisquare Spatial and Matern32 Temporal Basis Function	Exponential Spatial and Matern32 Temporal Basis Function	Matern32 Spatial and Matern32 Temporal Basis Function	Matern32 Spatial and Exponential Temporal Basis Function
LOOCV	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
Aksaray	12.1065	12.2447	11.9755	12.3789	12.3344	12.5441
Umraniye	9.3920	9.7646	9.7105	9.0167	9.2403	9.3756
Basiktas	13.2308	13.3406	13.3271	14.8686	14.7408	14.2934
Esenler	10.5712	10.2615	10.5730	9.7162	9.7065	9.7316
Esenyurt	18.4579	18.7876	18.9342	20.5982	20.0287	19.4824
Kartal	10.7937	10.9611	10.7347	10.4567	10.4169	10.8187
Kodikoy	13.2275	13.0958	13.3579	11.8482	11.9578	12.2006
Saryeri	18.4394	19.1571	19.9308	18.9089	19.6090	18.7359
Sile	17.0695	15.9573	15.5948	15.3947	15.3134	15.6900
Silivri	11.4963	12.0427	13.1218	9.9174	11.3160	10.9611
Sirinevler	8.4078	8.4574	9.0209	8.4999	8.6039	7.8630
Sultangazi	17.0238	17.2307	17.6058	18.9261	18.9885	18.4150
Basakeshir	17.8019	18.2404	18.5482	18.5216	19.0364	18.5397
Average	13.6937	13.8109	14.0335	13.7732	13.9456	13.7424

Nine random days were chosen from between two-month study period, as illustrated in Figure 1. Generally, the PM_{10} concentrations are stable, but on March 20, 2023, several areas experienced significantly higher levels. These included regions north of Aksaray and Esenler, as well as Basaksehir, Sultangazi, and Sariyer. Similarly, elevated levels were observed near the Umraniye and Kartal air quality monitoring stations in the southeast. The noticeable concentrations were also recorded between Sile and Esenyurt. On March 25, 2023, similar concentration patterns were observed in the same areas mentioned earlier, but the levels were lower than those recorded on March 20, 2023.

On the remaining days, there were no noticeable fluctuations in PM_{10} concentration, though small increases were observed in the previously mentioned areas. A clear connection between the presence of smartphone users and PM_{10} concentration was not observed, even when applying the same model fitting procedure without considering the covariate as the presence of smartphone users. There are several possible reasons for this. First, the two-month study period may be too short for spatial-temporal modeling, as it limits the capture of long-term trends and seasonal patterns. Second, this study relied on smartphone data from 443,786 users, which includes a low sampling rate as coordinates are recorded at irregular timestamps, and approximately 0.44 million is a small sample size relative to the broader population of Istanbul. To investigate the general relationship, the correlation between the total presence of smartphone users and PM_{10} concentration is computed as shown in Figure 2.

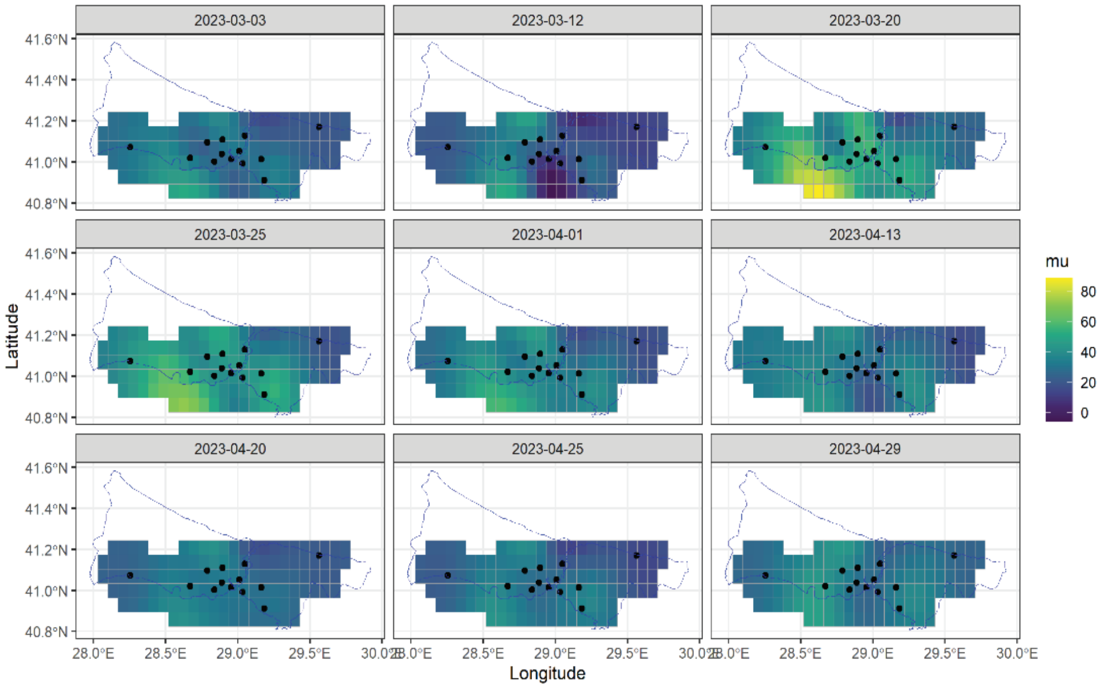

Figure 1: Prediction of average PM_{10} for randomly selected days

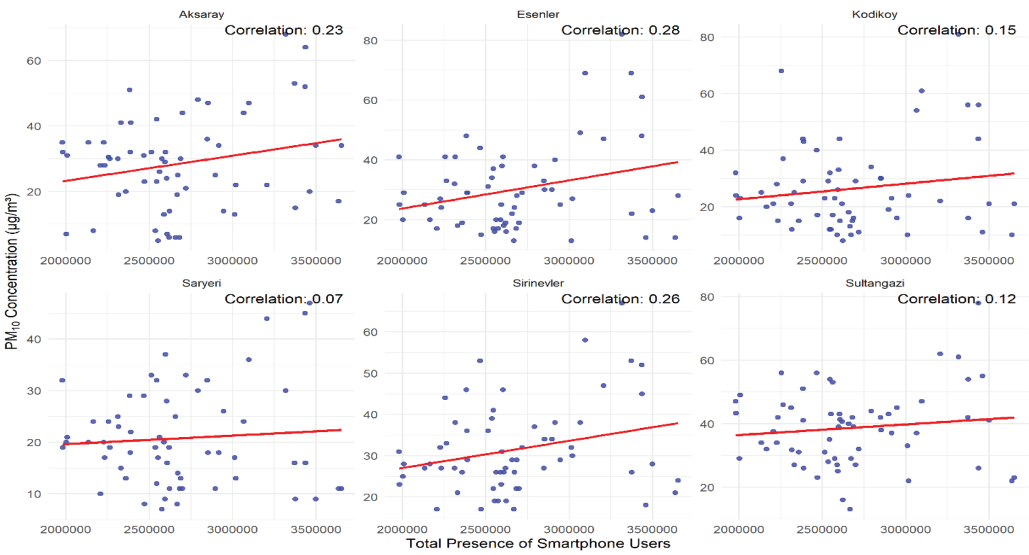

Figure 2: Scatter plot of total presence of smartphone users and PM_{10} concentration

From Figure 2, all air quality monitoring stations are located in the central area of Istanbul. It is observed that there is an overall very low correlation between the total presence of smartphone users and PM_{10} concentrations of all air quality stations. The correlation is 0.23,0.28, 0.15, 0.07, 0.26, and 0.12 for air quality stations such as Aksaray, Esenler, Kodikoy, Saryeri, Srinelver, and Sultangazi respectively. Overall, it is concluded that there is no correlation found between the PM_{10} concentrations and the total presence of smartphone users. It can be noticed that the total presence of smartphone users ranges from 0.2 million to about 0.36 million, showing daily variations in their presence across Istanbul. This variation could be due to factors like breaking in trajectory recording or smartphone users being temporarily outside the city. Future research would focus on the improvement of this FRK model by using longer-term historical air quality data and detailed individual mobility patterns with higher sampling rates. This approach would provide a clearer understanding of the relationship between air quality and human mobility. All statistical analysis was completed by using the R (version 4.3.2).

Acknowledgment

This contribution has been developed for the project "Study of mobile phone siGNals for the evalUation of the interconnections between mobility and the environment in Lombardy (SIGNUM)" CUP: F53D23010910001- PRIN 2022 PNRR M4C2 - financed by the European Union ˆa Next Generation EU (DD MUR n. 1409 del 14/09/2022)

References

Bucak, U., Arslan, T., Demirel, H., Balın, A. (2021). Analysis of strategies to reduce air pollution from vessels: A Case for the Strait of Istanbul. *Journal of Eta Maritime Science*, *9*(1), pp. 22–30. https://doi.org/10.4274/jems.2021.19327.

Finazzi, F., Paci, L. (2019). Quantifying personal exposure to air pollution from smartphone-based location data. *Biometrics*, *75*(4), 1356–1366. https://doi.org/10.1111/biom.13100.

Finazzi, F. (2016). The earthquake network project: Toward a crowdsourced smartphone-based earthquake early warning system. *Bulletin of the Seismological Society of America*, 106, pp. 1088–1099

Nguyen, H., Cressie, N., Braverman, A. (2012). Spatial statistical data fusion for remote sensing applications. *Journal of the American Statistical Association*, *107*(499), pp. 1004–1018. https://doi.org/10.1080/01621459.2012.694717.

Robinson, T. P., Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. *Computers and Electronics in Agriculture*, *50*(2), pp. 97–108. https://doi.org/10.1016/j.compag.2005.07.003.

Yurtseven, E., Vehid, S., Bosat, M., Köksal, S., Yurtseven, C. N. (2018). Assessment of ambient air pollution in Istanbul during. *Iran Journal Public Health,* 47 (8). http://ijph.tums.ac.ir.

Zammit-Mangion, A., Cressie, N. (2017). *FRK: An R Package for Spatial and Spatio-Temporal Prediction with Large Datasets*. http://arxiv.org/abs/1705.08105.

Zhu, Y., Xie, J., Huang, F., Cao, L. (2020). The mediating effect of air quality on the association between human mobility and COVID-19 infection in China. *Environmental Research*, *189*. https://doi.org/10.1016/j.envres.2020.109911.