MOX-Report No. 76/2023

The impact of public transport on the diffusion of COVID-19 pandemie in Lombardy during 2020

Ieva, F.; Galliani, G.; Secchi, P.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

The impact of public transport on the diffusion of COVID-19 pandemic in Lombardy during 2020

Greta Galliani[1], Piercesare Secchi[1], Francesca Ieva[1,2]

[1] MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy

[2] HDS, Health Data Science Center, Human Technopole, Milan, Italy

Abstract

In 2020, the COVID-19 pandemic has impacted the world, affecting health, economy, education, and social behavior. Much concern was raised about the role of mobility in the diffusion of the disease, with particular attention to public transport. Indeed, understanding the relationship between mobility and the pandemic is key for developing effective public health interventions and policy decisions.

In this work, we aim to understand how mobility, and more specifically mobility by public transport, has affected the diffusion of the pandemic at the regional scale. We focus our attention on Lombardy, the most populated Italian region severely hit by the pandemic in 2020. We explore static mobility data provided by Regione Lombardia, the regional service district, and dynamic mobility data provided by Trenord, a railway operator which serves Lombardy and neighboring areas. We develop an inventive pipeline for the dynamic estimation of Origin-Destination matrices obtained from tickets and passenger counts. This allows us to spot potential triggers in pandemic diffusion enhanced by the concept of proximity induced by mobility. We also develop a novel perspective for assessing the relationship between mobility and overall mortality based upon a functional approach combined with a spatial correlation analysis aimed at identifying the diversified effects on mortality in small geographical areas as a result of the restrictions on mobility introduced to contrast the pandemic.

Keywords: *COVID-19, mobility, OD matrix, spatial autocorrelation*

1 Introduction

The outbreak of the COVID-19 pandemic in 2020 has had a substantial impact on societies worldwide, posing significant challenges for public health systems and necessitating the implementation of unprecedented measures such as lockdowns, travel restrictions, and social distancing. Understanding the dynamics of epidemics and their relationship with various factors, including human mobility, is crucial for effective policymaking and outbreak response.[1] Mobility patterns play a critical role in the transmission and spread of infectious diseases as they determine the opportunities for contact and potential pathways for disseminating pathogens.[2] Analyzing the relationship between mobility flows and epidemic phenomena can provide valuable insights into the underlying mechanisms driving disease transmission and aid in formulating targeted intervention strategies.

Traditional approaches to study the link between mobility and epidemics have often relied on analyzing epidemiological data, such as the number of confirmed cases or mortality rates, together with variables or functions describing mobility flows as a whole, such as the total number of trips within the network, [3–5] indexes representing mobility for each area[6–9] or network characteristics like strength node centrality[10] or network efficiency.[11] While these approaches have proven helpful in understanding the broader dynamics of epidemics, they often overlook the spatial dimension and the intricate interplay between mobility and disease spread in the whole mobility network. The spatial dimension of epidemics refers to the geographic distribution of cases and the transmission patterns within and between different areas. Incorporating spatial analysis techniques into the study of epidemics can provide a more comprehensive understanding of how mobility influences disease propagation across different regions. These techniques are based on the definition of a nearness

notion which could be either based on the geographical closeness of such areas or induced by mobility, following the idea that the mobility-induced nearness is proportional to the number of movements between two areas. According to this new notion, two municipalities can be tightly related despite their distance.

In recent years, advancements in data collection and analysis techniques have opened up new possibilities for investigating the relationship between mobility and epidemics. The availability of large-scale mobility datasets, such as GPS traces from mobile devices, mobile phone network data, smart card data from public transportation systems, and geolocation data from social media platforms, has enabled researchers to capture fine-grained details of human movements.[12–15] These datasets provide a wealth of information that can be leveraged to quantify and model mobility patterns and explore their implications for disease transmission.[1]

This study proposes an innovative approach for analyzing the relationship between an epidemic phenomenon, specifically the COVID-19 pandemic, and mobility flows. We hypothesize that mortality rates should not be spatially correlated in times with no epidemic infections in the territory. On the contrary, in the case of perturbations caused by the pandemic (a spatially spreading phenomenon), we expect mortality rates to exhibit a positive spatial autocorrelation, with higher mortality rates in nearby areas. The motivation is that the epidemic spatially coordinates mortality, increasing it in nearby locations when it strikes. To validate our hypothesis, we analyze spatial autocorrelation in the mortality rates through global and local Moran indexes.[16,17] We focus on developing a notion of nearness induced by mobility flows as described by Origin-Destination (OD) matrices which represent the movement of individuals between different locations in a transportation network. We will compare this new notion of nearness with the geographical contiguity baseline, given that the epidemic spreads through proximity between individuals, and assess if mobility's proximity changes our analytical perspective.

To demonstrate the effectiveness of our approach, we apply it to two areas within the Lombardy region of Italy, which was severely affected by the COVID-19 pandemic. The first area covers the entire Lombardy region with fine spatial granularity, using static mobility data. The second area focuses on a smaller subset of Lombardy, specifically the provinces of Brescia, Bergamo, and Milano, and analyzes dynamic railway mobility flows. Dynamic data were obtained by data provided by Trenord, the local railway operator, about tickets and subscriptions sold together with data collected by the Automatic Passenger Counting (APC) system, now installed on a growing fraction of bus, metro and train rides worldwide and whose information is increasingly used to estimate mobility network dynamics.[18,19] By comparing the results obtained from these two areas, we can assess the robustness and generalizability of our approach across different spatial scales and types of mobility.

The objective of this study is threefold: (i) to investigate the relationship between mobility flows and the epidemic phenomenon by assessing spatial autocorrelation patterns induced by mobility; (ii) to compare the performance of mobility-based spatial weights with traditional contiguity-based spatial weights in capturing the spatial dynamics of the epidemic; and (iii) to explore the potential of dynamic mobility data (specifically railway mobility data) in enhancing our understanding of the temporal dynamics of the epidemic.

The rest of the paper is organized as follows: Section 2 describes the data used in the study and the statistical methodology related to spatial data analysis. Section 3 presents the obtained results in the two study areas, which are then discussed in Section 4. Finally, Section 5 summarizes our work's conclusion and spots relevant take-home messages.

2 Material and Methods

2.1 Data

To investigate the impact of mobility flows on the COVID-19 pandemic throughout 2020, we require data representing both epidemic and mobility aspects.

We employ mortality rates from all deaths' causes among individuals within each geographical area to model the epidemic response. The mortality rate $m_i^{[w]}$ is defined as the sum of deaths in area i between the current week w and a specific number of weeks before that week $w - e + 1$, divided by the population of area i:

$$m_i^{[w]} = \frac{\sum_{q=w-e+1}^{q=w} deaths_i^{[q]}}{population_i} \quad (1)$$

The daily death data for each day and each Italian municipality, as well as data about the population of each Italian municipality as of January 1, 2020, are publicly available from ISTAT, the Italian national institute of statistics. We experimented with different values for the aggregation parameter e in { 1,2,3,4 } trying to find a compromise between removing anomalies while avoiding over-smoothing in the data. Values of $w < 0$ encode the weeks of 2019, and values of $w > 52$ encode the weeks of 2021.

To derive mobility-based spatial descriptions, we need data about mobility flows in the form of OD matrices, which describe mobility in a transportation network. The cells t_{ij} of these matrices represent the number of trips starting from zone i and ending in zone j within a specific time frame. We employ two sources of such data: one from the Regione Lombardia (RL) Open Data program, which provides an OD matrix describing people's average movements during a workday in 2020, and another derived by us to capture the dynamic mobility changes during 2020 in a portion of the Trenord railway network.

The RL OD matrix, released in 2019, represents people's expected movements during a workday in 2020. [20] It provides estimates of movements divided into origin, destination, time slot, reason, and means of transport. It considers 1525 areas, with 1450 internal to Lombardy, while the others represent neighboring Italian provinces, non-neighboring Italian regions, neighboring Swiss districts, and foreign countries. The data account for 8 means of transport (car driver, car passenger, public transport by road, public transport by rail, motorcycle, bicycle, on foot, and others) and 5 reasons for movement (work, study, occasional, business, return). The spatial granularity mainly refers to municipalities, although some small towns are aggregated, and some large cities are distributed over multiple areas. We aggregate the time slots to represent mobility throughout the day and compute the sum of all movements for each reason and means of transport. We remove all OD couples concerning areas outside Lombardy and aggregate the zones referring to disaggregated large municipalities. The final OD matrix describes the total movements between 1360 zones representing the 1504 municipalities in Lombardy within a single workday.

Moreover, we derived a dynamic representation of weekly mobility flows within a specific area in Lombardy using data provided by Trenord. This type of data focuses on a peculiar kind of transport among those mentioned above, i.e., the public railway, being one of the most adopted by commuters, and allows to accurately represent and capture the fluctuations in mobility caused by the restrictions and behavior changes throughout the year.

The proposed method enhances the estimation of weekly OD matrices representing actual movements by train in a limited portion of the Trenord railway network. Two datasets were used for this purpose: one describing tickets and subscriptions purchased in December 2019 and throughout 2020, and the other reporting passengers boarding and dropping train rides for six train lines traveling within the provinces of Brescia, Bergamo, and Milano. The data from trains equipped with the APC system cover the 2020 period, except for July and August. As a result, we derived 41 OD matrices, each representing one week of 2020 within the available data period. Figure 1 provides a visual representation of the area covered by the Trenord network analyzed in our study, while Figure 2 graphically depicts the estimation pipeline.

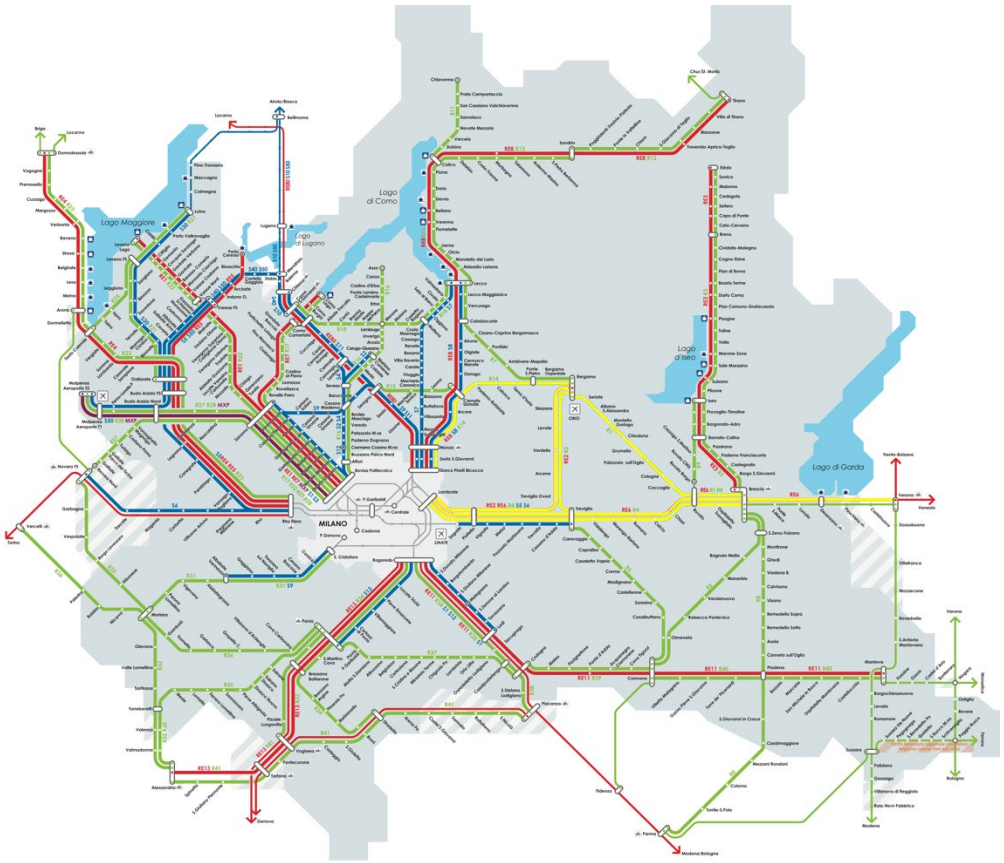

Figure 1: Map of the Trenord network. The train lines covered by data in our study are colored in yellow.

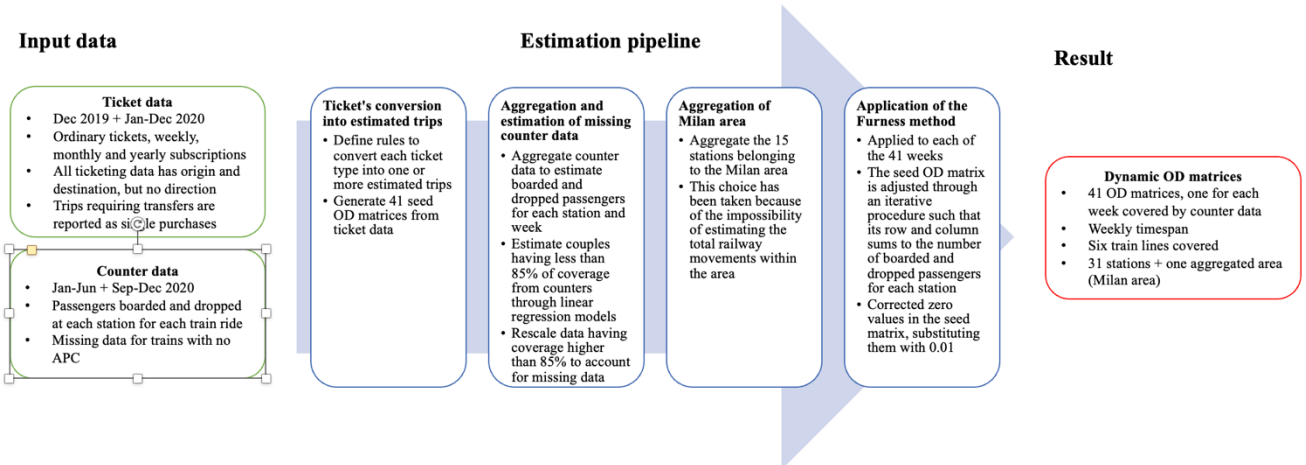

Figure 2: Estimation pipeline to obtain dynamical OD matrices taking in input ticket and counter data. The pipeline is based on the Furness method [21] for trip distribution modeling.

Notice that when estimating the Trenord OD matrices, we aggregated the 15 stations internal to the Milan and Monza provinces. This choice has been taken because we miss data about several train lines traveling within this area (lines S5, S6, S7, S8, S9, S11, RE8 and RE80 in Figure 1), making it impossible to represent the totality of movements by train. Since this would bias the mobility-based spatial weights we will define in the next station, we aggregate all the stations for which we do not have data about all the train lines traveling to or from other stations belonging to the selected portion of the network.

Both the static and dynamic representations will be employed in the spatial analysis described in Section 2.2 to define spatial weights for each area induced by mobility flows, enabling the assessment of local and spatial autocorrelation in the areal epidemic response.

To match the spatial granularity of the Trenord OD matrices, which refer to stations, with the granularity of the mortality response referring to municipalities, we utilize a dataset from ISTAT, providing distances in meters and minutes between Italian towns. The set of cities with the same closest station is considered the station basin for that particular station. Additionally, municipalities with a travel time exceeding 30 minutes to the nearest station are excluded since we evaluated that residents in those areas were more likely to opt for other means of transport than the train.

Following this procedure, we obtained the area shown in Figure 3, enclosing 261 towns divided into 28 station basins in the provinces of Brescia, Bergamo, and Milano, named "BreBeMi" from now on.

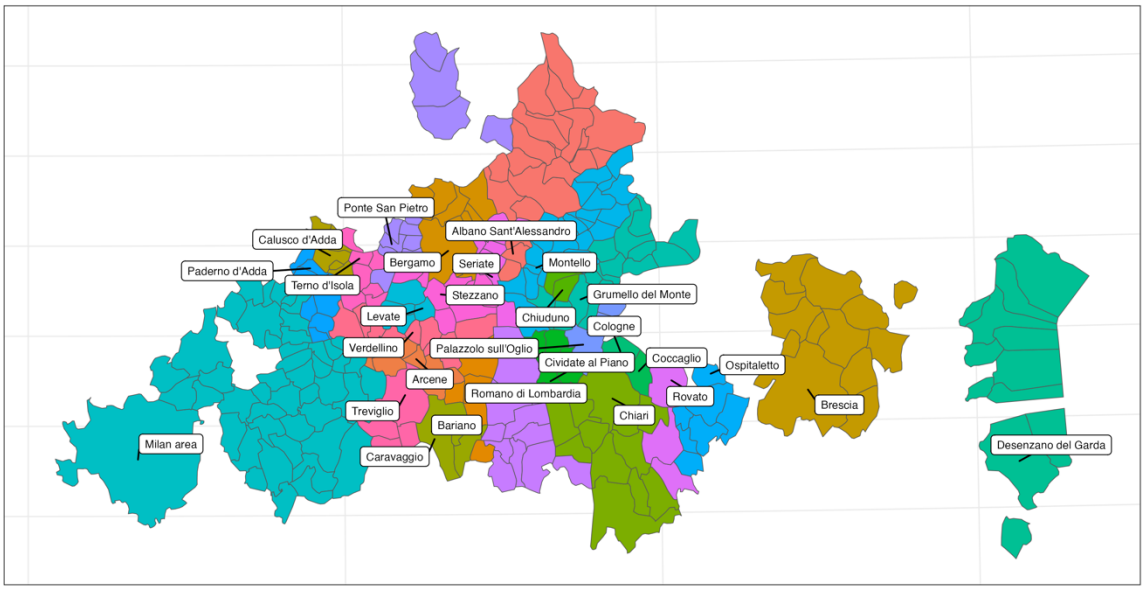

Figure 3: Map of the BreBeMi area induced by the Trenord mobility data, divided into station basins.

2.2 Spatial data analysis

To analyze the relationship between mobility flows and the epidemic response, we make use of spatial data analysis techniques[16,17,22,23]. We define a spatial metric describing closeness between the areas in the form of weight matrices induced by mobility and then identify global and local spatial autocorrelation in the areal epidemic feature (the mortality rates) according to this notion of proximity. Suitable spatial weights[22,23] quantify the relationship between each data point in a spatial dataset and its neighboring data points in the form of a matrix δ, with each element of the matrix δ_{ij} indicating the strength of the relationship between the two data points.

We define two spatial weight matrices in our analysis, one based on contiguity and the other derived from mobility:

1. **Contiguity-based spatial weights:** A purely geographical description is adopted: two areas are contiguous if they share a common border or vertex; in this case, we set $\delta_{ij} = 1$. Otherwise, we set $\delta_{ij} = 0$.
2. **Mobility-based spatial weights:** The mobility flows between two areas are used. Suppose we have an OD matrix describing, for each ordered couple of areas (i, j), the number of movements t_{ij} from i to j during a specific period. Then, we can define spatial weights in the given period as

$$\delta_{ij} = \frac{t_{ij}}{\sum_{j \neq i} t_{ij}}$$

The spatial weights are row-standardized, dividing t_{ij} for the total number of outgoing trips from i, $\sum_{j \neq i} t_{ij}$. This definition is asymmetrical since usually $t_{ij} \neq t_{ji}$. This kind of spatial weights, as opposed to contiguity-based ones, has rarely been applied in past studies mainly because of the difficulty in obtaining accurate mobility data and inferring spatial relations based on mobility flows.[24] An application can be found in Zhu et al.[25], where spatial weights are defined from mobility flows similarly and used to build spatial econometric models for the diffusion of COVID-19 in China.

Lastly, when dealing with dynamic mobility data (i.e., the Trenord OD matrices), a further parameter b varying in $\{1,2,3,4\}$ defines an aggregation of the mobility data between the current week w and a specific number of weeks before that week $w - b + 1$. More precisely:

$$\delta_{ij}^{[w]} = \frac{\sum_{q=w-\mathrm{b}+1}^{w} t_{ij}^{[q]}}{\sum_{j \neq i} \sum_{q=w-\mathrm{b}+1}^{w} t_{ij}^{[q]}}$$

This aggregation aims at smoothing fluctuations in the weekly mobility data.

In the analyses, we consider the spatial autocorrelation displayed by the weekly mortality rate $m_i^{[w]}$ according to the contiguity or mobility-induced spatial description δ. We employ global and local Moran indexes to analyze global and local spatial autocorrelation, hypothesizing that positive spatial autocorrelation reveals an ongoing epidemic phenomenon. Moran's index[16] is one of the most commonly used indicators for spatial autocorrelation. It is defined as:

$$I^{[w]} = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij} \left(m_i^{[w]} - \overline{m^{[w]}}\right)\left(m_j^{[w]} - \overline{m^{[w]}}\right)}{\sum_{k=1}^{n} \left(m_k^{[w]} - \overline{m^{[w]}}\right)^2} \qquad (2)$$

where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}$. In the case the dynamic Trenord OD matrices are used, a further parameter is introduced (lag in $\{1,2,\dots,10\}$), defining the delay between the mobility and mortality data. In this case, the Moran index becomes:

$$I^{[w]} = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}^{[w-lag]} \left(m_i^{[w]} - \overline{m^{[w]}}\right)\left(m_j^{[w]} - \overline{m^{[w]}}\right)}{\sum_{k=1}^{n} \left(m_k^{[w]} - \overline{m^{[w]}}\right)^2} \quad (3)$$

Figure 4 provides a visual representation of the three parameters introduced e, lag and b.

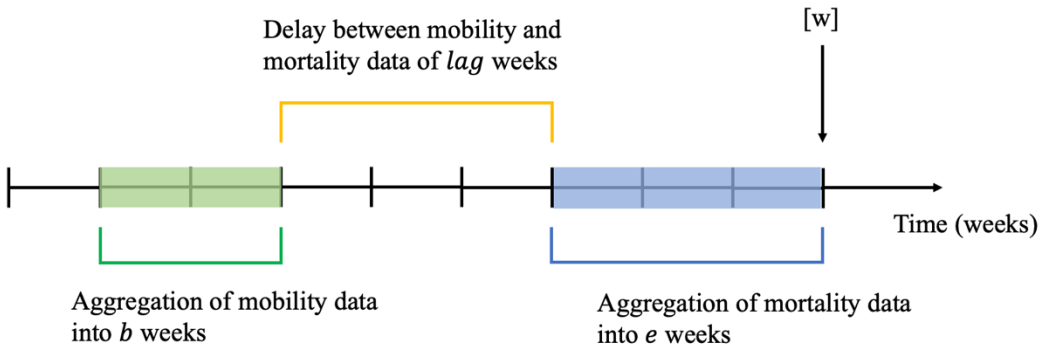

Figure 4: Graphical representation of the parameters e, lag and b.

We compute the global Moran indexes for each week of the mortality data and each dynamic mobility representation. We apply a permutational approach[26] to test for weeks showing positive spatial autocorrelation, adjusting the p-values for multiple testing through the False Discovery Rate procedure. Spanning w in $[1,\dots,52]$ (i.e., the entire 2020), we get a sequence of Moran indexes. We then analyze the sequence to spot periods showing positive spatial autocorrelation and interpret them in light of the epidemic development.

Finally, local Moran indexes[17] allow us to reveal spatial clusters (i.e., areas with very similar neighbors) and spatial outliers (i.e., areas with very different neighbors). Also in this case, starting from the traditional definition:

$$I_i^{[w]} = \frac{n}{S_0} \frac{\left(m_i^{[w]} - \overline{m^{[w]}}\right) \sum_{j=1}^{n} \delta_{ij} \left(m_j^{[w]} - \overline{m^{[w]}}\right)}{\sum_{k=1}^{n} \left(m_k^{[w]} - \overline{m^{[w]}}\right)^2} \quad (4)$$

The spatial weights δ_{ij} can be either static or dynamic, considering the lag value ($\delta_{ij}^{[w-lag]}$). We also test for significant values of the local indexes, adjusting the p-values through the False Discovery Rate procedure as in the global case. Index i varies in the number of areas defined from the spatial granularity, $[1,1360]$ for the RL data and $[1,28]$ for the Trenord one. Coupling the significance level with the interpreted Moran scatterplot,[22] we can identify spatial clusters and outliers. In particular, high-high spatial clusters highlight areas heavily affected in mortality whose neighboring areas also show higher mortality rates than average, indicating epidemic hotspots.

3 Results

In this Section, we show the results obtained by applying the spatial data analysis techniques in Section 2.2 to the two areas identified by the mobility datasets described in Section 2.1. First, we analyze the area of the Lombardy region, comparing the mobility spatial weights derived by the RL static OD matrix describing overall mobility in the area with the spatial weight matrix induced by contiguity. Then, we consider the BreBeMi area: we motivate the need for dynamic mobility data and reproduce the spatial analysis of mortality rates through the dynamic railway mobility spatial weights, comparing the results with the static railway RL mobility data in the same area.

For each of the two levels of analyses, we always consider the mortality rate $m_i^{[w]}$ accounting only for people aged 70 or more. This choice was taken based on the work done by Scimone et al. [27], which highlighted that this age class suffered the highest mortality risk caused by the pandemic and expressed the epidemic dynamics through 2020 in two waves of higher-than-average mortality rates. Considering the aggregation parameter in the mortality rates described by Equation (1), we select $e = 3$ as it provides a compromise between smoothing fluctuations and capturing weekly trends of increasing or decreasing mortality rates.

3.1 Analysis of Lombardy area

In this first level of analysis, we focus on the Lombardy area, considering the spatial granularity induced by the RL dataset, which roughly corresponds to municipalities. We compare two spatial weights: the static weight matrices defined from the OD RL dataset and the classical contiguity weights. We will refer to the RL mobility data with the term static overall mobility to highlight that data do not change in time and represent mobility from any transportation methods. We aim to identify weeks with positive spatial autocorrelation in mortality rates through global Moran indexes and to identify hotspots or influential areas through local Moran indexes, interpreting the indexes together with the pandemic development in 2020.

We compute global and local Moran indexes describing spatial autocorrelation in the mortality rates $m_i^{[w]}$ and for each week w of 2020 ($w \in [0,52]$) and each of the 1360 areas under investigation ($i \in [1,1360]$). Figure 5 shows the values of the global Moran indexes for each week of 2020, comparing mobility-based and contiguity-based spatial weights. The red points on the graph indicate a significantly positive spatial autocorrelation for the epidemic phenomenon described by $m_i^{[w]}$ in a given week (i.e., adjusted p-values smaller than 0.05).

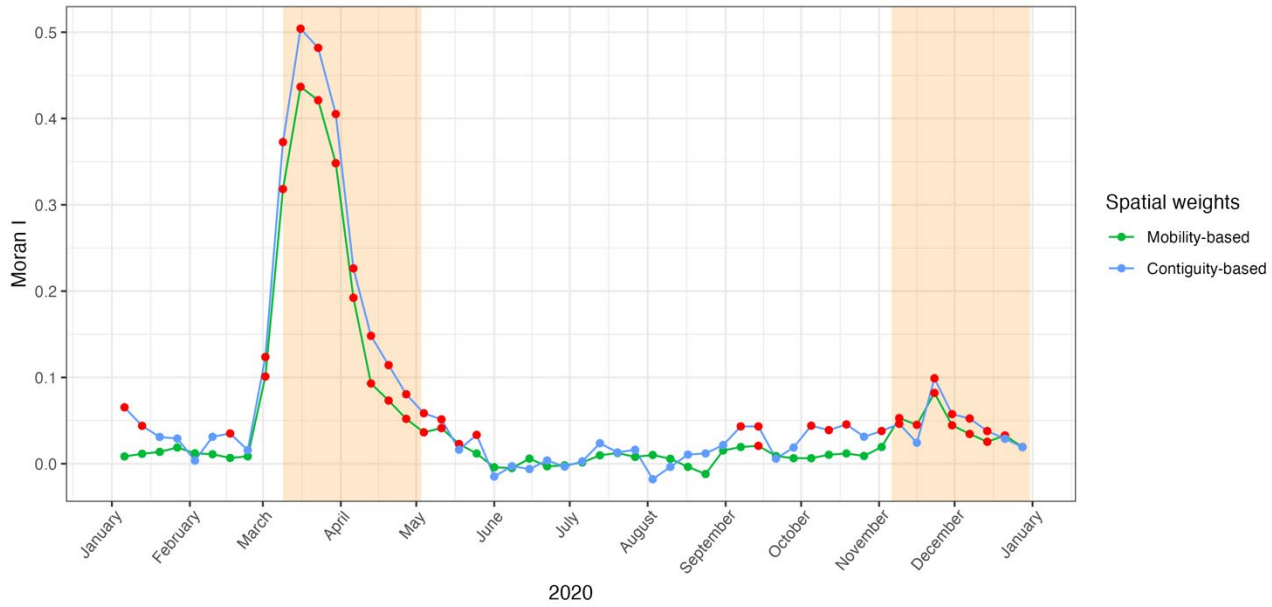

Figure 5: Global Moran indexes in the Lombardy area through 2020, comparing mobility-based and contiguity-based spatial weights. Red points indicate adjusted p-values < 0.05. Orange bands highlight the two lockdown periods in Lombardy.

The phenomenon exhibits a strong dependence on both spatial weights during the epidemic outbreak (March) and immediately after, with positive spatial autocorrelation observed from the beginning of March until the end of May. After May, neither spatial weights reveals significantly positive spatial autocorrelation in the mortality indexes. This coincides with the end of the national lockdown, suggesting that the substantial restrictions implemented during the lockdown period eliminated the spatial pattern of mortality. Comparing the two spatial weights, we observe that the corresponding Moran indexes are significantly greater than 0 in approximately the same weeks during the first wave. Both weights are nearly equivalent in revealing the epidemic phenomenon, but the Moran indexes induced by contiguity weights are greater than those generated by mobility weights.

Starting from the first week of November, we observe another period of positive spatial autocorrelation that continues until the end of the year, corresponding to the second wave period. There is a clear increase in the curve depicting the evolution of the global Moran index, although the curves are more oscillating and less pronounced than what is observed during the first wave. This trend persists, although it is arguable that the projected RL OD matrix no longer describes mobility trends after March. However, while the contiguity-based spatial weights flag a more extended period of positive spatial autocorrelation starting in October, the spatial weights derived from the RL OD matrix still highlight some positive spatial autocorrelation in the epidemic phenomenon. This could be interpreted as either suggesting that the projected movements may still provide valuable insights into mobility patterns or that during the second wave, mobility may no longer be the source of infections and the cause behind the positive spatial autocorrelation flagged by the mobility-based spatial weights is their similarity to other spatial weights effectively capturing the epidemic phenomenon.

The analysis of local Moran indexes allows us to highlight positive spatial clusters known as high-high areas, i.e., areas with mortality rates higher than average whose neighbors (according to the notion of nearness induced by mobility or contiguity spatial weights) also experience higher-than-average mortality rates. We focus on high-high areas because of their potential impact on the

spread of the pandemic. Figure 6 depicts the map representing significant areas revealed by the local Moran indexes in the second week of March, again comparing the mobility-based and contiguity spatial descriptions. Some red zones can be observed in the Val Seriana area, Codogno and its surroundings, and the Brescia province. These extensive high-high areas are visible in the maps from the beginning of March until mid-April, although the Codogno area disappears from the maps at the end of March, much earlier than the others. We notice that mobility-based spatial weights reveal larger clusters than contiguity-based weights throughout the first wave period: this suggests that mobility-based spatial weights provide insights into the dynamics of the epidemic phenomenon that are less captured by geographical metrics like contiguity. No extended spatial clusters are detected during the second wave period.

This first level of analysis indicates that mobility flows contain relevant information about epidemic dynamics in the local Lombardy area. However, despite the fine spatial granularity of the RL data, their static nature represents a significant limitation in our reasoning. This limitation led us to introduce dynamic OD data into the analysis.

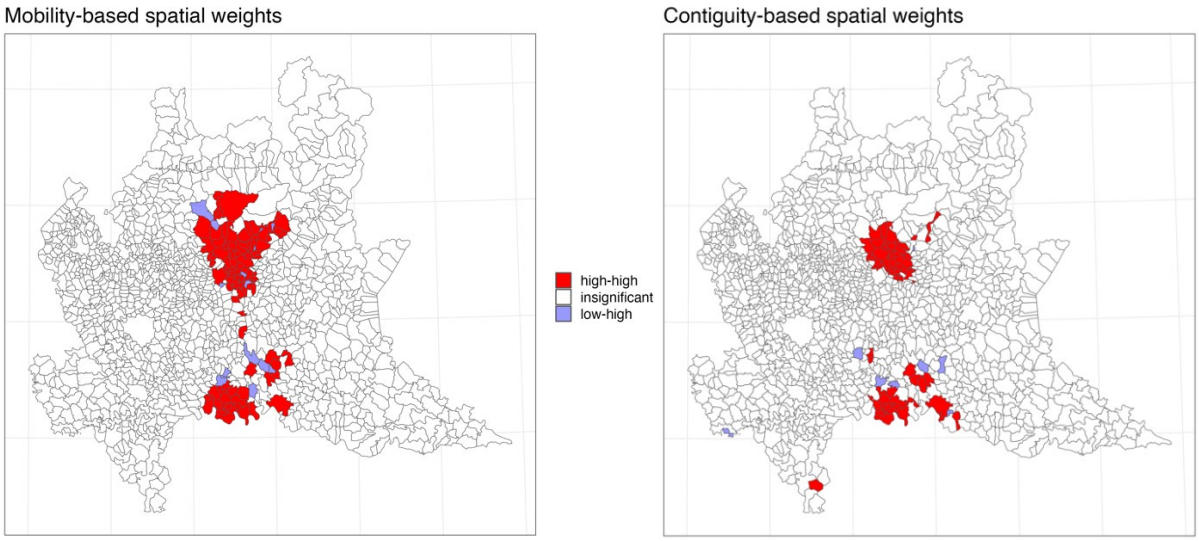

Figure 6: Spatial clusters and outliers identified in Lombardy during the 2nd week of March 2020 (from March 3 to March 15). The left panel shows the results induced by mobility-based spatial weights, while the right panel considers contiguity-based spatial weights.

3.2 Analysis of Brescia-Bergamo-Milano (BreBeMi) area

Having established the connection between local mobility flows in Lombardy and the evolution of mortality rates in 2020, our analysis expands to focus on a specific type of mobility profoundly affected by restrictions and subjected to behavioral changes after the pandemic: public railway mobility. To investigate this aspect, we employ dynamic OD matrices that depict railway mobility within a limited portion of the Trenord network. The Trenord data enables us to overcome the significant limitation of the RL data, which are static and do not capture variations in mobility flows over time. This limitation is crucial because restrictions and the fear of the disease caused significant fluctuations in mobility flows throughout 2020. [28] Therefore, we aim to extend our analysis by considering these mobility variations and investigating their impact on the epidemic phenomenon. To accomplish this, we turn to Trenord, a private company that operates the majority of railway mobility in Lombardy. We leverage their precise data to derive weekly OD matrices describing railway mobility flows within a limited area in Lombardy according to the pipeline reported in Figure 2.

Initially, we explore the relationship between the movements represented by the dynamic Trenord matrices and the static RL matrix, which portray average workday flows at the beginning of 2020. To this purpose, we select and aggregate the RL OD matrix to match the same area and spatial granularity of the Trenord data, considering only movements by train. We estimate linear regression models for each of the 41 weeks covered by the Trenord data, aiming to model each Trenord OD cell using the corresponding cell of the RL static data as a predictor. Figure 7 illustrates the R^2 coefficient of the regression models, while Figure 8 shows the regression coefficient's values.

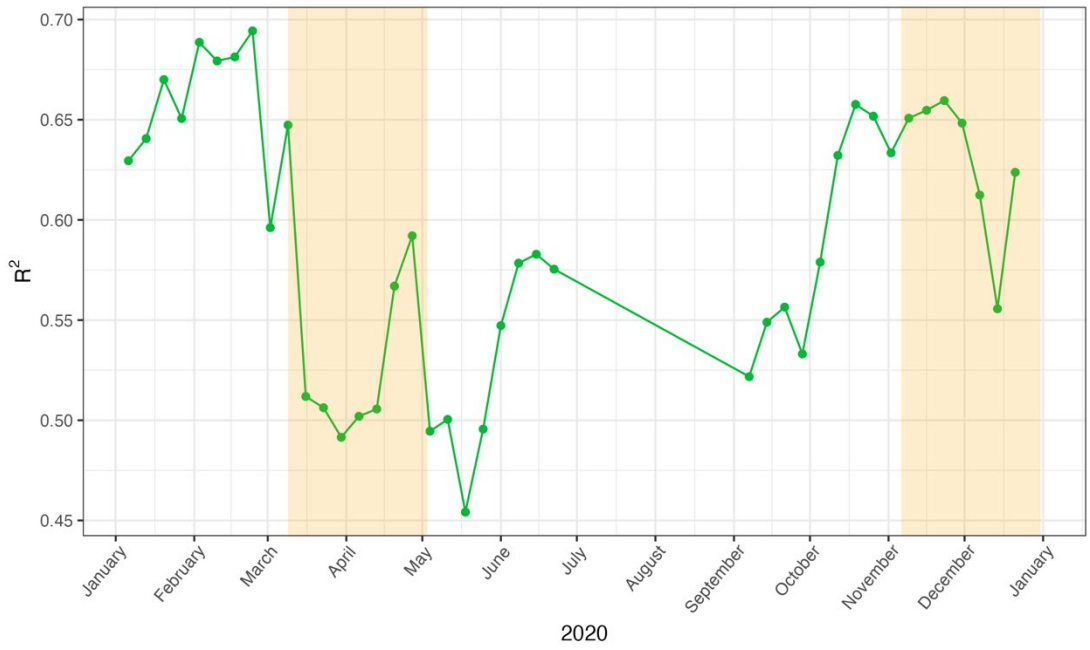

Figure 7: R^2 of the regression models predicting the dynamic Trenord OD cells through the corresponding static RL OD matrix cells. Lockdown periods are highlighted in orange.

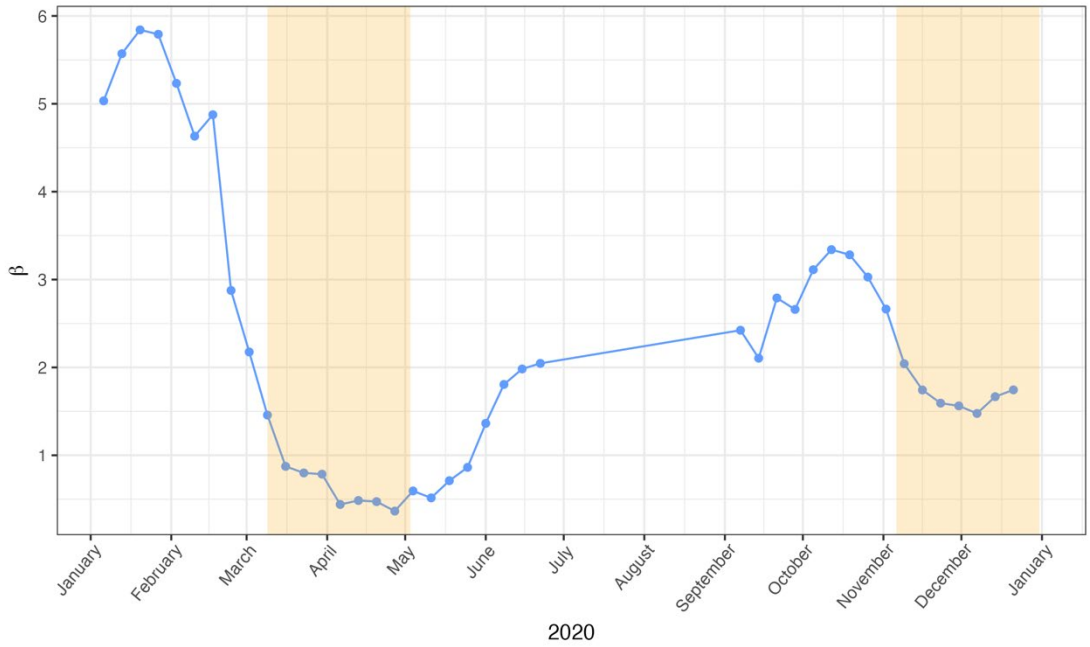

Figure 8: β regression coefficient of the models predicting the dynamic Trenord OD matrices through the corresponding static RL OD matrix cells. Lockdown periods are highlighted in orange.

The average value of the R^2 coefficient is 0.59, indicating a reasonably strong predictive power of the RL mobility data for Trenord movements. The R^2 index varies throughout 2020, exhibiting initially high values in January and February and lower ones during the first lockdown period (March to June). However, it is notable that the R^2 index returns to assume high values in the October, November and December months. This could indicate that after the beginning of October, people returned to use public railway transport in similar proportions to the face of the year, albeit the traffic volumes have decreased significantly compared to January and February, as seen in Figure 8. The close linear relationship between the two mobility data at the beginning of the year confirms the validity of our estimation procedure in deriving dynamic railway mobility flows closely resembling the reality described by the RL data. After this period, the changes in the R^2 index show the evolution of mobility and underscore the need for dynamic mobility data in analyzing the epidemic phenomenon. Consequently, we replicate the spatial analysis of the epidemic indicators within the area induced by the Trenord data, incorporating dynamic spatial weights. Notice that the RL data represent a single workday, while Trenord matrices refer to weeks. Thus, if the two data were perfectly coherent, we would expect $5 < \beta < 7$.

When working with dynamic spatial weights derived from the Trenord dynamic OD matrices, we introduce two additional parameters into the analysis, as described in Section 2.2: b and lag. After analyzing values of b within the range {1,2,3,4}, we determine that setting $b = 2$ minimizes anomalies in the weekly mobility trends while dynamically representing changes in railway flows. The lag parameter assumes particular significance, as it estimates the delay between the mobility phenomenon and its effect on the epidemic response as reflected in mortality rates. Consequently, we explore various values of lag, ranging from 1 to 10.

Figure 9 reports the curves obtained by computing the global Moran indexes for the weeks covered by the Trenord data and comparing them with the corresponding railway RL mobility data in the same area and spatial granularity.

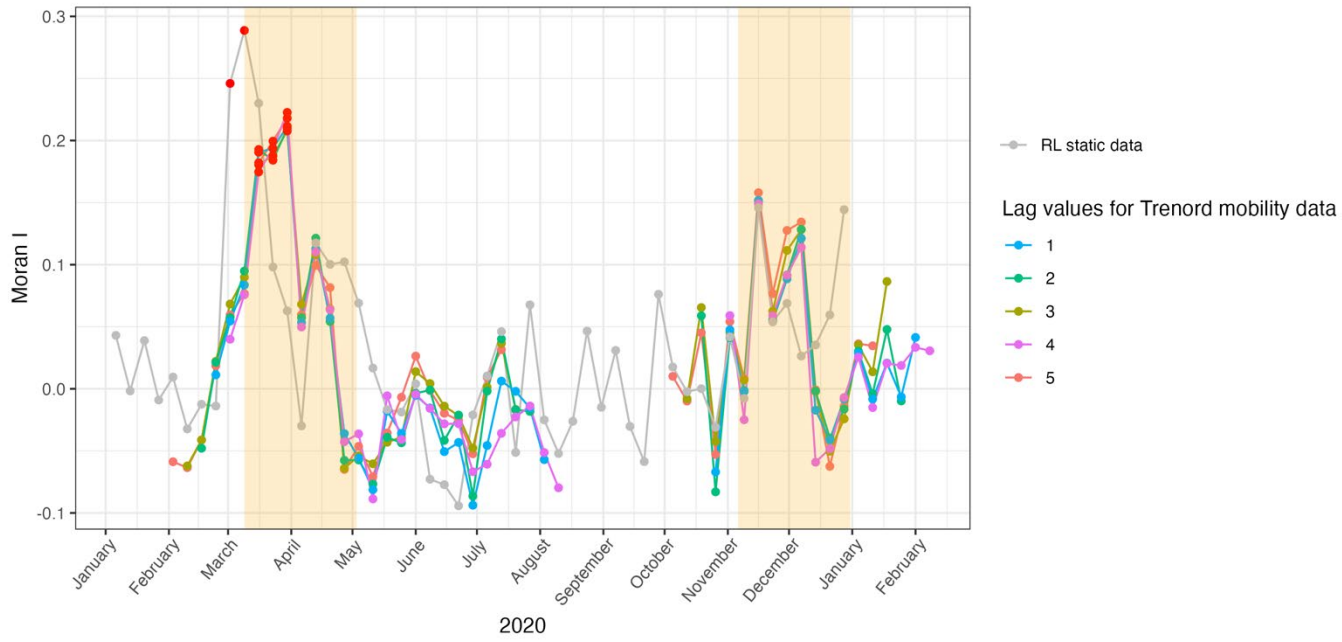

Figure 9: Global Moran indexes in the BreBeMi area, comparing two mobility data sources (Trenord and RL) and considering lag *in* {1,2,3,4,5} *for the dynamic Trenord data. Red points indicate adjusted p-values < 0.05. Lockdown periods are highlighted in orange.*

The figure reveals two peaks in the global Moran indexes values, indicated by both the RL and Trenord railway mobility-based spatial weights, albeit at different times. The index values decrease after the beginning of the first lockdown, highlighting its role in breaking the connection between mobility and the epidemic. In November, the curves rise again, even though we cannot refuse the hypothesis of spatial randomness with the permutational tests. Notice that when comparing the RL railway mobility-based spatial weights with the Trenord ones, we notice that while both the weights reveal peaks in the first wave period, the peaks happen at different times: the RL data reach the maximum value of the global Moran index in the second week of March and then drops after the third, while the Trenord weights show their peak in the last week of March, some weeks delayed than the RL data. The peak in the second wave period is reached for both spatial weights in the third week of November.

The parameter lag introduced for the dynamical Trenord data also influences the analysis. In Figure 9, we plot only the lag values revealing similar trends in the curves and highlighting the same weeks of positive spatial autocorrelation. For values of $lag \leq 5$, we always observe three weeks in March where the hypothesis of spatial randomness can not be refused. The curves' shapes look very similar in all these cases.

4 Discussion

In this paper, we proposed an innovative approach for analyzing the relationship between an epidemic phenomenon (COVID-19) as described by mortality rates and mobility flows. We aim to demonstrate that accurate and reliable mobility data are crucial in providing valuable insights to guide policymakers in making informed decisions regarding future lockdown and restriction policies.

This work assessed global and local spatial autocorrelation in an areal epidemic indicator using Moran indexes and retrieving spatial weights from OD matrices. We tested this approach in two areas within the Lombardy region: one covering the entire region with fine spatial granularity using static mobility data and the other focusing on a limited portion of Lombardy and analyzing dynamic railway mobility flows.

The developed pipeline is entirely general, as it establishes a connection between an areal epidemic indicator and OD matrices describing different forms of mobility in one or more transportation networks. By modifying the type of mobility data and epidemic indicators, we can analyze various aspects of the pandemic: while mortality rates of people aged 70 or older have proven to be representative in past studies, other areal epidemic data, such as positive swab rates, quarantined individuals, or occupancy of intensive care units, can be considered to represent different responses associated with an epidemic phenomenon. When analyzing pandemic evolution beyond 2020, it is essential to note that subsequent waves have primarily impacted peak infection rates rather than mortality rates. [10] Thus, it may be necessary to employ different epidemic indicators when examining later periods.

In the Lombardy area, we compared the static mobility spatial weights based on overall mobility described by the static RL data with classical contiguity-based ones. Both spatial weights identified positive spatial autocorrelation in the epidemic feature during similar periods corresponding to the first and second wave periods. However, contiguity-based spatial weights exhibited higher global Moran index values than mobility-based ones. The power of mobility-based spatial weights emerged in detecting spatial clusters, with mobility-based weights revealing larger positive spatial clusters during the first wave period. The role of the flagged high-high areas in epidemic diffusion requires further investigation to determine if these areas acted as starting locations for the spread

through their outbound mobility flows. The flagged high-high areas in both maps (obtained applying the mobility or contiguity based spatial weights) correspond to the Val Seriana area, Codogno and its surroundings, and some towns in Brescia province, already highlighted as areas heavily affected by the epidemic by other works in the field. [27] Analyzing the dynamic maps produced by the local Moran indexes according to the mobility or contiguity spatial weights, we notice how the Codogno area disappear from the map some weeks before the others, coherently with early mobility restrictions applied in this territory, which was placed under lockdown two weeks earlier than the rest of Lombardy. This shows our method's capability to detect phenomena consistent with known pandemic dynamics. We point out that the mobility-based spatial weights signal the non-significance of the Codogno area one week earlier than the contiguity-based ones.

In summary, the analysis of the Lombardy area concludes that mobility-based spatial weights can highlight ongoing epidemic phenomena through global Moran indexes and provide insights into local patterns that cannot be identified using classical contiguity-based weights.

Moreover, the spatial analysis highlighted the role of the first national lockdown in breaking the positive spatial autocorrelation. Indeed, the global Moran indexes decrease in this period. After a while, no positive spatial autocorrelation is detected in the mortality rates: the restrictions had a consequence in the spatial patterns of the mortality rates. Moreover, the high-high areas flagged in the local spatial analysis disappeared sometime after the beginning of the first lockdown. This is evidence confirming the hypothesis presented in the Introduction: pandemic outbreaks introduce positive spatial autocorrelation in the mortality rates, while when the epidemic subsides, the mortality rates return to being spatially uncorrelated.

Thus, we turned to the dynamic Trenord data. We developed a pipeline to construct dynamic OD matrices representing movements by train, selecting only a limited portion of the network (six train lines) because of the avability of these data. Despite the small area covered, these data allow us to describe actual mobility trends throughout 2020 and examine the role of a specific type of public transport heavily impacted by restrictions. Comparing the dynamic Trenord mobility data with the static RL data describing railway mobility in the same area and spatial granularity, we observed a strong relationship between the two early in the year ($R^2 \geq 0.65$ in January and February). This confirms the validity of the Trenord data, as they closely match the RL data in periods unaffected by pandemic-related disruptions. However, deviations between the two data sources emerged during the initial weeks of the pandemic outbreak and subsequent restrictions, supporting the adoption of dynamic data describing mobility after March 2020. As a result, we repeated the spatial autocorrelation analysis between mobility data and mortality rates in the restricted BreBeMi area, focusing solely on railway mobility. Moreover, dynamic mobility analysis allowed us to model the lag representing the delay in weeks between mobility flows and their influence on the epidemic phenomenon. Accurate estimates of this parameter are crucial for policymakers to evaluate the timeframe for mobility restrictions to show their impact on the pandemic response.

Another advantage of dynamic OD matrices lies in their ability to reveal the temporal evolution of mobility patterns, allowing us to identify shifts in travel behavior and potential spreading events in response to changes in the epidemiological situation. Traditional static OD matrices may not adequately capture human movement dynamics, especially during rapidly changing circumstances like an ongoing epidemic. Incorporating dynamic OD matrices into the analysis of epidemic dynamics opens avenues for future research and policy development. As the world faces recurrent outbreaks and potential future epidemics, monitoring human mobility through dynamic OD matrices can assist policymakers in making data-driven decisions on implementing targeted restrictions, contact tracing efforts, and vaccination campaigns. Moreover, the application of

dynamic OD matrices is not limited to COVID-19 alone; it can be extended to monitor the spread of other infectious diseases, enhancing our preparedness and response to future health crises.

Our analysis of the BreBeMi area revealed that mobility flows were still associated with two peaks in the values of the global Moran index, showing an effect of railway mobility flows on the epidemic phenomenon. However, there were noticeable differences when comparing the curves with those observed in the Lombardy area. In the BreBeMi area, the curves displayed oscillations and lower index values, with fewer weeks flagged for significant spatial autocorrelation than the corresponding dataset using Lombardy data. These observations can be partially attributed to the limited spatial coverage and granularity of the BreBeMi area, representing only a smaller portion of Lombardy. Both areas exhibited peaks in the Moran index values corresponding to the first and second wave periods, although they occurred at different times for the RL and Trenord railway mobility-based spatial weights: the peak in the first lockdown shown by the Trenord data was delayed compared to those computed according with the RL data. We noticed that we did not observe any lag between the contiguity and RL mobility-based spatial weights in the Lombardy area. Coupling these observations together, we could hypothesize that the RL dataset accurately captures commuter traffic, generating a nearness metric more similar to contiguity than the Trenord data. On the other hand, Trenord data better depicts occasional traffic and enhances the relevance of mobility to and from the Milan area, as emerged from our analysis of the Trenord data. This aspect may be investigated through the fusion of actual data with different further dynamic sources, like cellphone data collected by mobile operators or GPS positions.[15,29] Moreover, the BreBeMi area considered in the second level of analysis was heavily affected by the epidemic in terms of mortality during the first wave. Thus, the second wave was lightly perceived in the area, likely because the population at high risk had already significantly shrunk, and survivors had developed natural immunity due to the large circulation of the virus during the first wave, as suggested by Scimone et al.[27] This could explain why the permutation tests do not flag any weeks showing significantly positive spatial autocorrelation during the second wave period.

By analyzing the role of the lag parameter, we highlighted that values of $lag \leq 5$ gave evidence of positive spatial autocorrelation in the same three weeks of the first wave period and resulted in extremely similar curves. Our estimates of mobility's impact on pandemic flows within six weeks provides a starting point for future analyses on the matter, which are crucial for policymakers to put in place effective mobility restrictions. Moreover, our findings are coherent with other works in the field estimating the impact between reductions in mobility and their effect on mortality rates and new confirmed cases. [9,10]

There are limitations to our work that have to be acknowledged. First, our spatial analysis pipeline relies on the assumption that we have reliable OD matrices accurately describing mobility flows. While the OD matrices obtained from Trenord data contain valuable dynamic information, our analysis revealed some estimation problems, such as potential overestimation of flows at certain stations, like the Milan area and Verona. Despite these limitations, the data align with reality-induced principles, showing intense fluxes between connected stations and fluctuations corresponding to lockdown periods or easing of restrictions. Improvements in estimation procedures can be explored in future research, considering other methods in trip distribution models such as gravity models. [30] This is not an easy point to accomplish, especially in the dynamic case where tickets and passenger counters over impose numbers of simplification. Additionally, the limited data coverage of dynamic mobility data and the coarse spatial granularity pose further limitations. However, a strength of the pipeline developed is the possibility to scale it to cover the entire Trenord network or to adapt it to other transportation networks, provided that seed OD matrices and estimates of trips beginning and ending in each zone are available. The relevance of the estimation pipeline extends beyond our research, as dynamic OD matrices provide a real-time representation of

mobility patterns that could be useful to address themes related to transportation or assessing socio-economic trends.[31]

5 Conclusion

In this study, we have introduced an innovative approach to analyze the relationship between mobility flows and epidemic phenomena. By employing mobility-based spatial weights induced by static and dynamic OD matrices, we have successfully identified spatial autocorrelation patterns using global and local Moran indexes, offering valuable insights into the patterns and impacts of mobility on epidemics. We have applied our methodology to two distinct areas within the Lombardy region: one encompassing the entire region with fine spatial granularity, albeit relying on static mobility data, and the other focusing on a smaller area comprising the provinces of Brescia, Bergamo, and Milano, where we compared static and dynamic railway mobility flows.

The results of our analysis have revealed a robust relationship between mobility-based spatial weights and mortality rates in the examined areas. The curves of global Moran indexes exhibit two distinct peaks of positive spatial autocorrelation, consistently corresponding to the two epidemic wave periods, regardless of the mobility description (static or dynamic) or the specific focus on railway or overall mobility. Notably, within the Lombardy area, mobility-based spatial weights have allowed us to identify larger hotspot areas compared to classical contiguity weights, underscoring the potential of spatial descriptions based on mobility to uncover hidden patterns within the epidemic phenomenon. Furthermore, the comparison between Trenord dynamic mobility data and RL static data in the BreBeMi area has provided distinct perspectives, revealing temporal shifts in the peaks of global spatial autocorrelation compared to static data. Additionally, our investigation into the lag between dynamic mobility flows and their impact on mortality rates has led us to argue that such lag is less than six weeks.

Despite the inherent limitations and challenges encountered throughout our study, our approach presents promising avenues for future research aimed at comprehending and managing the intricate relationship between mobility flows and epidemics. Investigating further methods for deriving accurate mobility data is crucial, as this would significantly enhance our understanding of epidemic dynamics. In this framework, we developed a pipeline to exploit data collected by Trenord, the local railway operator, and derive dynamic OD matrices, highlighting some benefits and limitations of our procedure.

Lastly, by modifying the type of mobility data and epidemic indicators, our spatial analysis pipeline can be readily adapted to explore various aspects of the pandemic, offering indispensable information for policymakers and researchers. This valuable knowledge can better equip us to tackle future epidemics, ensuring we are prepared to address and mitigate their impact effectively.

6 Acknowledgments

This research is a part of the CHANCE (Chronic diseases management after the CoViD19 epidemic trigger. Capturing data, generating evidence, suggesting actions for health protection) project. The project has been supported by Fondazione Cariplo, grant n° 2020-4238

We thank Trenord for the collaboration and for sharing the data used in this work. In particular, we thank Dr. Marta Galvani and Dr. Giovanni Chiodi for their support and insightful suggestions.

The authors acknowledge the support by MUR, grant Dipartimento di Eccellenza 2023-2027.

7 References

1. Hu T, Wang S, She B, et al. Human mobility data in the COVID-19 pandemic: characteristics, applications, and challenges. *Int J Digit Earth*. 2021;14(9):1126-1147. doi:10.1080/17538947.2021.1952324
2. Kraemer MUG, Golding N, Bisanzio D, et al. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource-poor settings. *Sci Rep.* 2019;9(1):5151. doi:10.1038/s41598-019-41192-3.
3. Pieroni V, Facchini A, Riccaboni M. COVID-19 vaccination and unemployment risk: lessons from the Italian crisis. *Sci Rep*. 2021;11(1):18538. doi:https://doi.org/10.1038/s41598-021-97462-6
4. Kartal MT, Depren Ö, Kiliç Depren S. The relationship between mobility and COVID-19 pandemic: Daily evidence from an emerging country by causality analysis. *Transp Res Interdiscip Perspect*. 2021;10:100366. doi:https://doi.org/10.1016/j.trip.2021.100366
5. Kishore N, Kahn R, Martinez PP, Salazar PM De, Mahmud AS, Buckee CO. Lockdowns result in changes in human mobility which may impact the epidemiologic dynamics of SARS-CoV-2. *Sci Rep*. 2021;11(1):6995. doi:10.1038/s41598-021-86297-w
6. Hu S, Xiong C, Yang M, Younes H, Luo W, Zhang L. A big-data driven approach to analyzing and modeling human mobility trend under non-pharmaceutical interventions during COVID-19 pandemic. *Transp Res Part C Emerg Technol*. 2021;124:102955. doi:https://doi.org/10.1016/j.trc.2020.102955
7. Vahedi B, Karimzadeh M, Zoraghein H. Spatiotemporal prediction of COVID-19 cases using inter- and intra-county proxies of human interactions. *Nat Commun*. 2021;12(1):6440. doi:10.1038/s41467-021-22108-2
8. Ilin C, Annan-Phan S, Tai XH, Mehra S, Hsiang S, Blumenstock JE. Public mobility data enables COVID-19 forecasting and management at local and global scales. *Sci Rep*. 2021;11(1):13531. doi:10.1038/s41598-021-92824-3
9. Iacus SM, Santamaria C, Sermi F, Spyratos S, Tarchi D, Vespe M. Human mobility and COVID-19 initial dynamics. *Nonlinear Dyn*. 2020;101(3):1901-1919. doi:10.1007/s11071-020-05854-6
10. Mazzola V, Bonaccorsi G, Ieva F, Secchi P. The effects of mobility restrictions on public health: a functional data analysis for Italy over the years 2020 and 2021. Poster presented at: *SEAS Conference*; June 21-23th, 2023; Ancona.
11. Bonaccorsi G, Pierri F, Cinelli M, et al. Economic and social consequences of human mobility restrictions under COVID-19. *Proc Natl Acad Sci U S A*. 2020;117(27):15530-15535. doi:10.1073/pnas.2007658117
12. Alsger AA, Mesbah M, Ferreira L, Safi H. Use of Smart Card Fare Data to Estimate Public Transport Origin–Destination Matrix. *Transp Res Rec*. 2015;2535(1):88-96. doi:10.3141/2535-10

13. Willumsen LG. *Estimation of an OD Matrix from Traffic Counts–A Review.* Institute of Transport Studies, University of Leeds; 1978.
14. Moreira-Matias L, Gama J, Ferreira M, Mendes-Moreira J, Damas L. Time-evolving O-D matrix estimation using high-speed GPS data streams. *Expert Syst Appl*. 2016;44:275-288. doi:https://doi.org/10.1016/j.eswa.2015.08.048
15. Manfredini F, Pucci P, Secchi P, Tagliolato P, Vantini S, Vitelli V. Treelet Decomposition of Mobile Phone Data for Deriving City Usage and Mobility Pattern in the Milan Urban Region. In: Paganoni AM, Secchi P, editors. *Advances in Complex Data Modeling and Computational Methods in Statistics.* Springer International Publishing; 2015. p. 133-147. doi:10.1007/978-3-319-11149-0_9.
16. Moran PAP. The Interpretation of Statistical Maps. *J R Stat Soc Ser*. 1948;10(2):243-251. doi:https://doi.org/10.1111/j.2517-6161.1948.tb00012.x
17. Anselin L. Local Indicators of Spatial Association - LISA. *Geogr Anal*. 1995;27(2):93-115. doi:https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
18. McCord MR, Mishalani RG, Goel P, Strohl B. Iterative Proportional Fitting Procedure to Determine Bus Route Passenger Origin–Destination Flows. *Transp Res Rec*. 2010;2145(1):59-65. doi:10.3141/2145-07
19. Torti A, Galvani M, Urbano V, et al. Analysing transportation system reliability: the case study of the metro system of Milan. Technical Report 84, *MOX-Report No. 84/2021*. 2021. https://www.mate.polimi.it/biblioteca/add/qmox/84-2021.pdf. Accessed April 13, 2023.
20. Regione Lombardia. Matrice OD2020 - Passeggeri. Published 2019. Accessed July 17, 2023. https://www.dati.lombardia.it/Mobilit-e-trasporti/Matrice-OD2020-Passeggeri/hyqr-mpe2
21. Evans AW. Some properties of trip distribution methods. *Transportation Res*. 1970;4:19-36. doi:https://doi.org/10.1016/0041-1647(70)90072-9
22. Anselin L, Syabri I, Kho Y. GeoDa: An Introduction to Spatial Data Analysis. *Geogr Anal*. 2006;38(1):5-22.
23. Bivand RS, Pebesma E, Gòmez-Rubio V. *Applied Spatial Data Analysis with R*. Springer; 2013.
24. Ermagun A, Levinson D. An Introduction to the Network Weight Matrix: Introduction to the Network Weight Matrix. *Geogr Anal*. 2017;50. doi:10.1111/gean.12134
25. Zhu P, Li J, Hou Y. Applying a Population Flow-Based Spatial Weight Matrix in Spatial Econometric Models: Conceptual Framework and Application to COVID-19 Transmission Analysis. *Ann Am Assoc Geogr*. 2022;112(8):2266-2286. doi:10.1080/24694452.2022.2060791
26. Cliff AD, Ord JK. *Spatial Processes: Models and Applications*. Pion Limited; 1981.
27. Scimone R, Menafoglio A, Sangalli LM, Secchi P. A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities. *Spat Stat*. 2022;49:100541. doi:https://doi.org/10.1016/j.spasta.2021.100541
28. Smolyak A, Bonaccorsi G, Flori A, Pammolli F, Havlin S. Effects of mobility restrictions during COVID19 in Italy. *Sci Rep*. 2021;11(1):21783. doi:10.1038/s41598-021-01076-x
29. Bayir MA, Demirbas M, Eagle N. Discovering spatiotemporal mobility profiles of cellphone users. In: *2009 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks & Workshops*. 15-19 June 2009:1-9. doi:10.1109/WOWMOM.2009.5282489
30. Wilson AG. The Use of Entropy Maximising Models, in the Theory of Trip Distribution, Mode Split and Route Split. *J Transp Econ Policy*. 1969;3(1):108-126.
31. Toch E, Lerner B, Ben-Zion E, Ben-Gal I. Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowl Inf Syst*. 2019;58(3):501-523. doi:10.1007/s10115-018-1186-x

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

75/2023 Archetti, A.; Ieva, F.; Matteucci, M.
Scaling survival analysis in healthcare with federated survival forests: A comparative study on heart failure and breast cancer genomics

74/2023 Pidò, S.; Pinoli, P.; Crovari, P.; Ieva, F.; Garzotto, F.; Ceri, S.
Ask Your Data—Supporting Data Science Processes by Combining AutoML and Conversational Interfaces

71/2023 Conni, G.; Piccardo, S.; Perotto, S.; Porta, G.M.; Icardi, M.
HiPhome: HIgh order Projection-based HOMogEnisation for advection diffusion reaction problems

70/2023 Ragni, A.; Ippolito, D.; Masci, C.
Assessing the Impact of Hybrid Teaching on Students' Academic Performance via Multilevel Propensity Score-based techniques

69/2023 Ferro, N.; Micheletti, S.; Parolini, N.; Perotto, S.; Verani, M.; Antonietti, P. F.
Level set-fitted polytopal meshes with application to structural topology optimization

68/2023 Vitullo, P.; Colombo, A.; Franco, N.R.; Manzoni, A.; Zunino, P.
Nonlinear model order reduction for problems with microstructure using mesh informed neural networks

67/2023 Conti, P.; Guo, M.; Manzoni, A.; Hesthaven, J.S.
Multi-fidelity surrogate modeling using long short-term memory networks

66/2023 Fresca, S.; Gobat, G.; Fedeli, P.; Frangi, A.; Manzoni, A.
Deep learning-based reduced order models for the real-time simulation of the nonlinear dynamics of microstructures

65/2023 Gatti, F.; de Falco, C.; Perotto, S.; Formaggia, L.; Pastor, M.
A scalable well-balanced numerical scheme for the modelling of two-phase shallow granular landslide consolidation

Gatti, F.; de Falco, C.; Perotto, S.; Formaggia, L.; Pastor, M.
A scalable well-balanced numerical scheme for the modeling of two-phase shallow granular landslide consolidation