POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

A spatial analysis of railway mobility and its relationship with pandemic spread in Lombardia during 2020

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Greta Galliani

Advisor: Prof. Francesca Ieva

Co-advisor: Prof. Piercesare Secchi

Academic year: 2021-2022

---

1. Introduction

In the early months of 2020, COVID-19 swept the world, causing massive disruptive effects in our societies. The virus circulated in Italy quickly, and the government was forced to take drastic measures to slow its transmission. These measures resulted in a sharp reduction in people leaving their homes and travelling within cities and across regions. Consequently, travel demand fell, and usage of private vehicles and public transportation dropped dramatically. Local public transit was much affected by restrictions, fear of the disease, and lifestyle changes. However, while there are multiple shreds of evidence about the role of mobility in the spread of the disease, the public transport impact still has to be assessed as past research came to different (and sometimes contrasting) conclusions.

This work aims to assess the role of mobility in disease diffusion, focusing on public railway transport in Lombardia during the year 2020. Investigating this matter is key to take active interventions in future outbreaks, not only of COVID-19 but of other diseases. The hypothesis that public transport did not have such a prevalent role in epidemic spread should be analysed in detail. If confirmed, it calls for a review of some restrictions on its usage for future outbreaks.

To reach this goal, we investigate the link between epidemic and mobility data by applying spatial data analysis techniques. We select one areal epidemic indicator (mortality rates) and define a spatial description derived from mobility flows. We aim to detect periods of positive spatial autocorrelation in the epidemic feature according to the mobility-based spatial description to show the influence of mobility in the epidemic phenomenon. Then, we identify areas showing clustered behaviour (similar mortality rates in nearby areas, according to the mobility-induced notion of nearby) to highlight epidemic hotspots. We need mobility data accurately describing trends in 2020 concerning both overall mobility (i.e., by all means of transportation) and railway one. To estimate dynamic mobility data describing weekly railway mobility flows, we define a pipeline taking in input data provided by Trenord, the local railway company.

This paper is organised as follows: Section 2 presents the epidemic and mobility data used in the study, Section 3 develops the pipeline to estimate the dynamic mobility data needed to address our research goal, and Section 4 details the framework to analyse the relationship between mobility flows and epidemic spread and applies it, first in the entire Lombardia region and then in the limited area covered by Trenord data. Finally, Section 5 discusses the implications and conclusions of our findings.

2. Data

In analogy to what was done in [1], we model the epidemic response with the weekly mortality rate m of people aged 70 years or older.

To derive mobility-based spatial descriptions, we need data about mobility flows in the form of Origin-Destination (OD) matrices (i.e., matrices describing mobility in a transportation network, whose cells t_{ij} represent the number of trips starting from zone i and ending in zone j in a specific time frame). We first consider the movements of the OD matrix released by the Regione Lombardia (RL) Open Data program. The advantage of this dataset is its description of a wide area (Lombardia) and fine spatial granularity (municipalities), dividing movements by reason and means of transportation. However, the disadvantage is its static description derived from projections of past data. Thus, it is not a reliable description of mobility trends after the COVID-19 outbreak since it is not based on data collected in 2020.

The necessity of deriving real dynamic mobility data led us to develop a pipeline to estimate weekly OD matrices representing actual movements by train in a limited portion of the Trenord railway network. We were provided with two datasets to reach this goal, describing tickets purchased in 2020 and passengers boarding and dropping train rides for six train lines collected through trains equipped with the Automatic Passenger Counting (APC) system.

3. Estimation of Origin-Destination Matrices

We now detail the pipeline to estimate the Trenord dynamic OD matrices based on the Furness method for trip distribution modelling.

3.1. Furness method for trip distribution modelling

Trip distribution modelling predicts the number of trips between origins and destinations in a transportation network. The Furness method [2] is a technique used in this framework.

Fixing the time unit, we start with an origin and destination survey producing a seed matrix. In our case, we derive the seed matrix from ticket data. The cells of the seed matrix t^*_{ij} represent the number of survey-estimated trips beginning in zone i and ending in zone j, while $q_i = \sum_{j=1}^{J} t^*_{ij}$ is the number of trips beginning in zone i and $b_j = \sum_{i=1}^{I} t^*_{ij}$ is the number of trips ending at zone j. Since the seed matrix is generated through a survey, its row and column totals q_i and b_j do not generally equal the estimates of the trips starting and ending in each zone. Let estimates of the actual number of trips beginning in each zone be $p_1, \ldots, p_I$ and let $a_1, \ldots, a_J$ be the estimates for the number of trips ending in each zone. In our application, we derive estimates of real trips from the APC passenger count data.

The trip distribution problem is to derive from matrix t^*_{ij} a forecasting matrix t_{ij} whose row and column totals are respectively $p_1, \ldots, p_I$ and $a_1, \ldots, a_J$. To derive matrix t_{ij}, we need to iteratively find constants by which to multiply the original matrix t^*_{ij} elements.

The Furness method provides an answer to the trip distribution problem. At each passage, a matrix $t^{(n)}_{ij}$ is obtained by multiplying the previous matrix $t^{(n-1)}_{ij}$ by a suitable constant $x^{(n)}_{ij}$. It can be proven that the limiting matrix $t_{ij} = \lim_{n\to\infty} t^{(n)}_{ij}$ satisfies the required properties.

3.2. Estimation pipeline

We now define a four-step pipeline to estimate movements by train in 8 months (37 weeks) of 2020. We consider a limited portion of the Trenord railway network covering six train lines and 46 stations.

1. **Tickets' conversion into estimated trips:** we introduced assumptions to translate each ticket and subscription type into one or more ticket-estimated trips in 37 weekly seed OD matrices.
2. **Estimation of missing counter data:** we developed a model to estimate the number of boarded and dropped passengers for each station and week. The model uses data about the partial passenger counts and then corrects them to account for train rides

with missing counter data through a combination of linear regression models and a rescaling procedure.

3. **Aggregation of the Milan area:** because we do not possess data about all the train lines moving in the area surrounding Milan and because of the absence of a considerable number of tickets internal to the city of Milan, we aggregate the 15 stations of this area into a single zone.
4. **Application of Furness method:** finally, we apply the Furness method to the 37 ticket seed OD matrices and margin vectors of boarded and dropped passengers. We use a procedure to correct the cells having no estimated trips from the ticket conversion process based on a posteriori binomial test since the Furness method can not correct zero values.

The final estimated OD matrices after Furness show coherency with some reality-induced principles, displaying many movements around major centres and on paths belonging to the same train line, revealing decreasing trips during the two lockdowns and increases in periods of lesser restrictions. The margin errors, which measure the concordance between the estimated matrices and the margin vectors, are low (in the order of units) for all the matrices.

The purpose of this kind of mobility data in the following analysis is twofold: they describe actual mobility trends through 2020 and allow us to consider the role of a specific kind of public transport believed to be a primary carrier in epidemic diffusion and strongly impacted by restrictions.

4. Spatial analysis of mobility and epidemics

To address our research goal, we define a spatial description based on mobility data and employ global and local Moran indexes to analyse global and local spatial autocorrelation in mortality rates. The underlying hypothesis is that no spatial autocorrelation in the mortality rates would exist without an epidemic phenomenon. We consider two areas, first the entire Lombardia region and then the area covered by the Trenord data derived in the previous Section. Moreover, we compare the mobility-based spatial description with a purely geographical one in the form of zones' contiguity.

4.1. Spatial analysis methods

We define a novel type of spatial weights to infer a spatial description derived from mobility data. Starting from an OD matrix describing the number of movements t_{ij} from area i to area j in a specific time frame, we define spatial weights δ_{ij} as

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

Thanks to the mobility-derived spatial description, we can test for positive spatial autocorrelation through weekly global Moran indexes [3] computed considering the weekly mortality rate m:

$$I = \frac{n}{\sum_{i,j=1}^{n} \delta_{ij}} \frac{\sum_{i,j=1}^{n} \delta_{ij}(m_i - \bar{m})(m_j - \bar{m})}{\sum_{k=1}^{n}(m_k - \bar{m})^2}$$

Where $\bar{m}$ is the mean of the weekly mortality rates and n is the number of data areas.

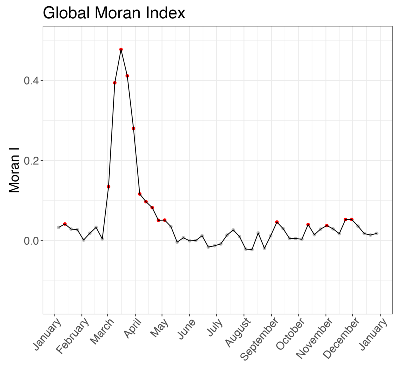
While global Moran indexes capture the degree of geographical clustering, we employ local Moran indexes to detect spatial clusters - observations with very similar neighbours - and spatial outliers - observations with very different neighbours. In this framework, Anselin [4] defines the local Moran index as:

$$I_i = \frac{n}{\sum_{i,j=1}^{n} \delta_{ij}} \frac{(m_i - \bar{m}) \sum_{j=1}^{n} \delta_{ij}(m_j - \bar{m})}{\sum_{k=1}^{n}(m_k - \bar{m})^2}$$

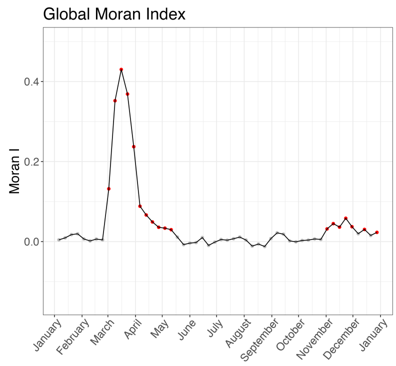
These indexes, coupled with the values of the mortality rates $(m_i - \bar{m})$ in the area and the spatially lagged values $\sum_{j=1}^{n} \delta_{ij}(m_j - \bar{m})$ in nearby areas, allow us to classify the nature of spatial autocorrelation into four categories. Positive spatial autocorrelation corresponds to high-high or low-low spatial clusters (similar values in nearby locations). In contrast, negative spatial autocorrelation (dissimilar values at neighbouring locations) identifies high-low and low-high spatial outliers.

4.2. Analysis of Lombardia area

To evaluate the role of mobility in epidemic diffusion, we first consider the RL mobility data released in 2019. While this kind of data does not offer information about movements that actually happened, it describes a wide area (all of Lombardia) at a fine spatial granularity.



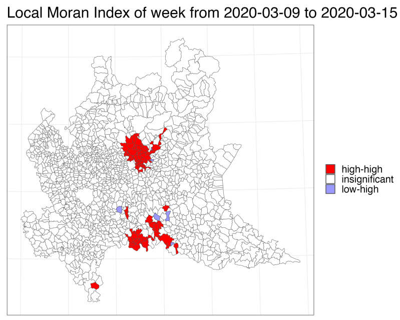
(a) Contiguity-based spatial weights



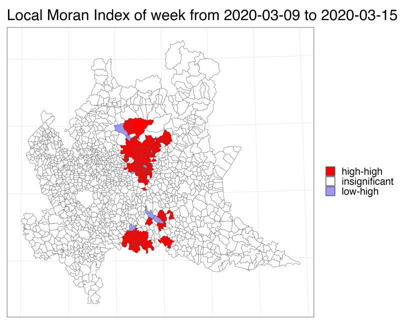
(b) Mobility-based spatial weights

Figure 1: Global Moran indexes at Lombardia level based on RL mobility data

We employ global and local Moran indexes to investigate the spatial association between the weekly mortality rates for each week of 2020 and area and a spatial description based on mobility, compared with a purely geographical one based on contiguity, where zones influence each other if they share a border or vertex. Figure 1 compares the global Moran indexes computed with the two kinds of spatial weights. The mobility-derived spatial description shows a period of significantly positive spatial autocorrelation in the epidemic phenomenon corresponding to the first wave in the same period as contiguity-based spatial weights. Moreover, the same mobility-based weights still show positive autocorrelation during the second wave period, even if the projected RL data are no more reliable to provide an accurate mobility description after the end of February. We can also notice the effect of the first national lockdown in extinguishing the link between mobility and contiguity spatial descriptions and the epidemic. Moreover, we observed no spatial autocorrelation in periods outside the first and second wave outbreaks. We can thus confirm our initial assumption: periods of no exponential epidemic phenomena do not display global or local spatial patterns.



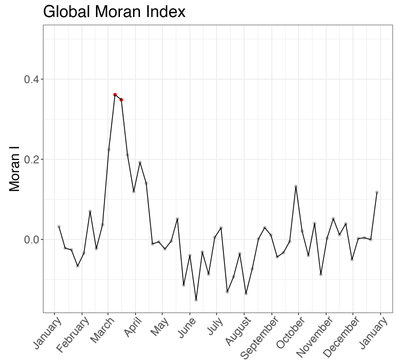
(a) Contiguity-based spatial weights



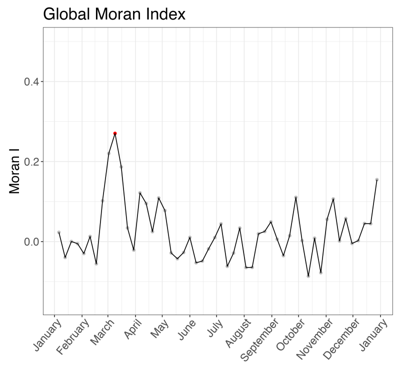
(b) Mobility-based spatial weights

Figure 2: Spatial clusters and outliers identified in the third week of March

However, the higher power in describing the epidemic of mobility weights compared to contiguity ones comes out in detecting spatial clusters through local Moran indexes. During the first wave, the mobility weights always reveal more extensive high-high areas (i.e., areas with higher-than-average values of the mortality rates whose neighbours also show higher-than-average values) compared to contiguity-based spatial weights. Figure 2 compares the local analysis derived by the two spatial weights in the third week of March. We detect high-high areas from March 2 to April 5, and the zones correspond to regions known to be strongly hit by the disease, such as the surroundings of Codogno or the Val Seriana area. The red zones disappear sometime after the implementation of the first lockdown. No spatial cluster is detected during the second wave period, suggesting that the epidemic may have spread homogeneously during the second outbreak without a characterisation in hotspots.



(a) Overall mobility-based spatial weights



(b) Railway mobility-based spatial weights

Figure 3: Global Moran indexes at BreBeMi level based on RL mobility data

4.3. Analysis of BreBeMi area

Because of the characteristics of the Trenord data, we repeat the spatial analysis in a limited area of Lombardia (named BreBeMi area because the majority of its territory belongs to the provinces of Brescia, Bergamo and Milano) and spatial granularity based on 28 areas identified by the stations. First, we retrace the spatial analysis of the RL mobility data, comparing spatial weights based on overall mobility with railway mobility. We notice how the results are much less interpretable in this area and granularity than those referring to the Lombardia area. However, we notice positive spatial autocorrelation in the spatial mobility description through global Moran indexes. Overall mobility identifies two weeks of March (during the epidemic outbreak) of positive spatial autocorrelation, while railway mobility identifies only one week of the two as significant. Thus, we can still infer a connection between mobility and the epidemic during the first wave, which loosens during the first lockdown. Figure 3 compares global Moran indexes computed according to overall mobility spatial weights and railway mobility ones.

We then consider the available Trenord estimated OD matrices describing real movements by train in the BreBeMi area through 2020 and repeat the spatial analysis. In this case, the spatial weights vary weekly, so we consider a lag in the interval [0,10] between spatial weights and mortality rates. In this framework, we have to underline that we miss Trenord data describing January and February's mobility, which is believed to have had the strongest impact on the first epidemic spread.

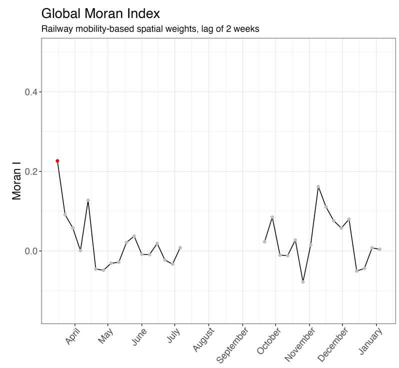


Figure 4: Global Moran indexes at BreBeMi level based on Trenord mobility data

The global Moran indexes show only one week with significantly positive spatial autocorrelation, corresponding to week 11 (in the mid of March) with the mobility-based spatial weights of week 9. Figure 4 shows the global Moran indexes computed through the available Trenord mobility data, with a two-week lag between mobility weights and the weekly mortality rates. No interesting area is detected in the local spatial analysis, likely because of the limited area and wide spatial granularity, which make the identification of significant local Moran indexes challenging.

5. Conclusion

This work was motivated by the interest in analysing mobility data to assess the role of mobility in the pandemic diffusion through 2020, focusing on evaluating if a particular kind of public transport (public railway transport) had a prevalent role in the COVID-19 spread. We analysed two mobility datasets, one provided by

Regione Lombardia and the other derived by Trenord's ticket and passenger count datasets, to assess if spatial descriptions based on mobility could identify positive spatial autocorrelation in mortality rates and detect spatial clusters and outliers.

To derive the Trenord dynamic OD matrices, we defined a pipeline taking in input tickets bought and partial passenger count data. The four-step pipeline presents some innovative features compared to other works in trip distribution modelling, like the conversion of tickets into estimated trips and the model developed to correct partial count data. The final estimated OD matrices demonstrate coherence with reality-induced principles and attain low margin errors. They provide a valuable tool for understanding actual mobility trends in 2020 and the impact of public railway transport on epidemic diffusion. The pipeline could be used in other areas and networks to estimate public transportation movements.

Coming to our research goal, we found that a strong relationship exists between mobility flows and the epidemic spread in Lombardia during 2020. Indeed, all our analyses through global Moran indexes highlighted periods of significantly positive spatial autocorrelation in mortality rates, according to all the mobility-based spatial descriptions employed. Moreover, we observed that mobility-based spatial weights could identify larger high-high spatial clusters than the contiguity-based spatial description in the Lombardia area during the first wave period. These areas warrant further investigation into their potential role in the epidemic's spread.

Then, we compared overall and railway mobility as described by the static RL OD matrix or by the Trenord-derived dynamical ones. We observed that railway mobility-based spatial weights show positive spatial autocorrelation with the mortality rates in two weeks of March, the period of strongest epidemic spread. Coupling these findings with the analysis of overall mobility-based spatial weights from RL data in the BreBeMi area, which identified the same weeks as showing significantly positive spatial autocorrelation, we can assert that railway mobility does have a link with epidemic diffusion. Still, this link is weaker than the one shown by overall mobility because of the lower global Moran indexes than the overall mobility case. There is no evidence of railway mobility being a more prominent disease carrier than overall mobility in any period of 2020. Indeed, we never found a positive spatial autocorrelation with the railway mobility-based spatial description in periods of no correlation with overall mobility.

Limitations of our methodology include the limited focus on the Lombardia and BreBeMi areas and the consideration of only two mobility data sources, Trenord and Regione Lombardia. The missing Trenord data describing January and February mobility could be recovered to deepen the analysis of dynamical railway mobility flows. Further investigation is needed to assess if mobility-based spatial descriptions could describe the epidemic diffusion's dynamic, assessing if two nearby areas' mortality rates influence each other through time-series causality tests. If this hypothesis is confirmed, our analysis might be the starting point for developing decision tools for policymakers to react to epidemic phenomena. Again, the role of different kinds of mobility, related or not to public transport, should be considered for future analysis to assess their contribution to the pandemic development.

6. Acknowledgements

This research is a part of the CHANCE project. We thank Trenord for the data provided.

References

[1] R. Scimone et al. A look at the spatio-temporal mortality patterns in Italy during the COVID-19 pandemic through the lens of mortality densities. *Spatial Statistics*, 49:100541, 2022.

[2] A. W. Evans. Some properties of trip distribution methods. *Transportation Research*, 4:19–36, 1970.

[3] P. A. P. Moran. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, 10(2):243–251, 1948.

[4] L. Anselin. Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27(2):93–115, 1995.