



European
Commission

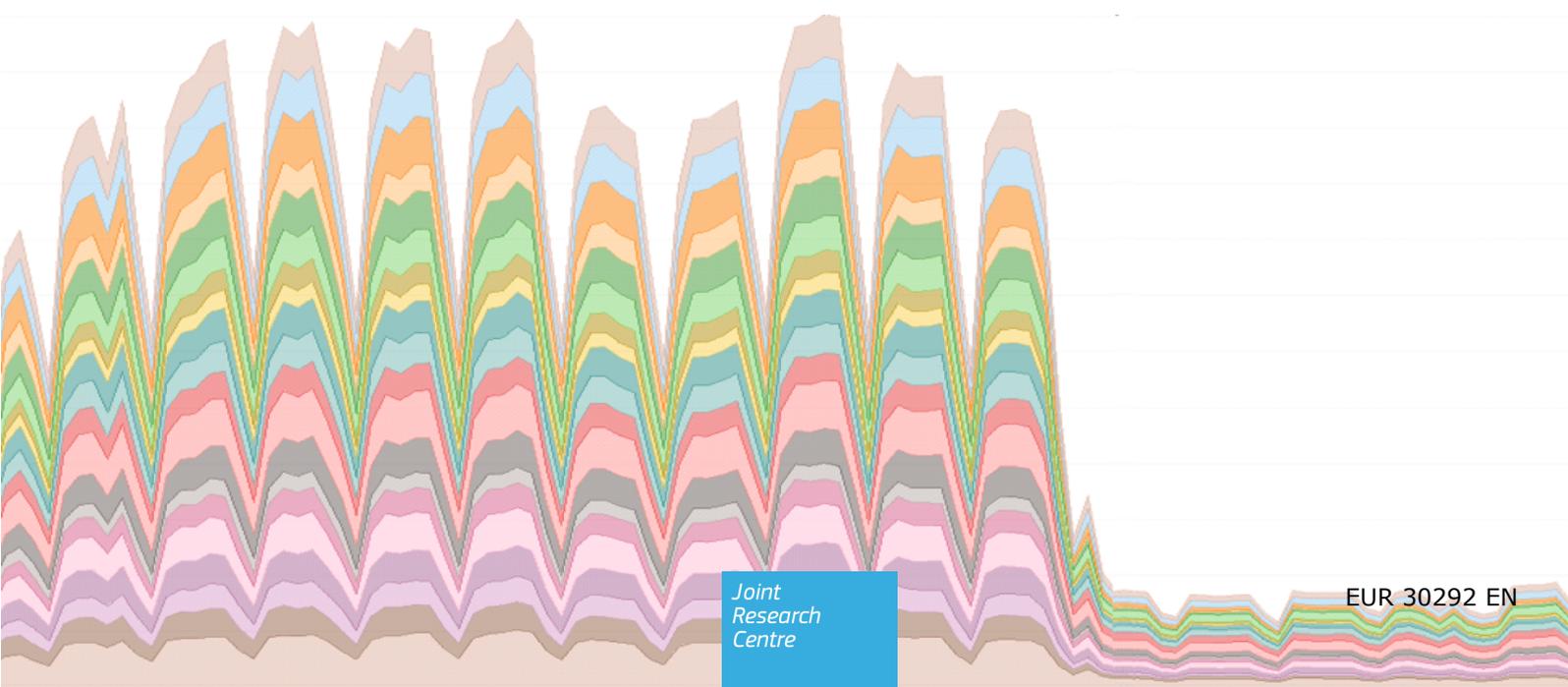
JRC TECHNICAL REPORTS

How human mobility explains the initial spread of COVID-19

A European Regional
Analysis

Iacus, S. M., Santamaria, C., Sermi, F.,
Spyratos, S., Tarchi, D., Vespe M.

2020



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact Information

Name: Stefano Maria Iacus

Address: Joint Research Centre, Via Enrico Fermi 2749, TP 26B, 21027 Ispra (VA), Italy

E-mail: stefano.iacus@ec.europa.eu

Tel.: +39 0332 786088

EU Science Hub

<https://ec.europa.eu/jrc>

JRC121300

EUR 30292 EN

PDF ISBN 978-92-76-20430-5 ISSN 1831-9424 doi:10.2760/61847

Luxembourg: Publications Office of the European Union, 2020

© European Union, 2020



The reuse policy of the European Commission is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Reuse is authorised, provided the source of the document is acknowledged and its original meaning or message is not distorted. The European Commission shall not be liable for any consequence stemming from the reuse. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2020

How to cite this report: Iacus, S., Santamaria, C., Sermi, F., Spyrtatos, S., Tarchi, D. and Vespe, M., How human mobility explains the initial spread of COVID-19, EUR 30292 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-20430-5, doi:10.2760/61847, JRC121300.

Contents

- Acknowledgements 1
- Abstract 2
- 1 Introduction 3
- 2 Mobile Positioning Data and Mobility Indicators 4
 - 2.1 Origin-Destination Matrix 4
 - 2.2 Connectivity Matrix 4
- 3 Human mobility explains the early spread of CODIV-19 in France..... 5
 - 3.1 Internal mobility matters the most..... 8
 - 3.2 Correlation network analysis 10
- 4 Human mobility and CODIV-19 spread in Italy..... 12
- 5 Human mobility and IgG antibody testing in Spain 15
- 6 Caveats of the study 15
- 7 Discussion 17
- References 19

Acknowledgements

The authors acknowledge the support of European MNOs (among which A1 Telekom Austria Group, Altice Portugal, Deutsche Telekom, Orange, Proximus, TIM Telecom Italia, Telefonica, Telenor, Telia Company and Vodafone) in providing access to aggregate and anonymised data, an invaluable contribution to the initiative.

The authors would also like to acknowledge the GSMA¹, colleagues from DG CONNECT² for their support and colleagues from Eurostat³ and ECDC⁴ for their input in drafting the data request.

Finally, the authors would also like to acknowledge the support from JRC colleagues, and in particular the E3 Unit, for setting up a secure environment, a dedicated *Secure Platform for Epidemiological Analysis and Research* (SPEAR) enabling the transfer, host and process of the data provided by the MNOs; as well as the E6 Unit (the Dynamic Data Hub team) for their valuable support in setting up the data lake.

Authors

Stefano Maria Iacus, Carlos Santamaria, Francesco Sermi, Spyridon Spyratos, Dario Tarchi, Michele Vespe

¹GSMA is the GSM Association of Mobile Network Operators.

²DG Connect: The Directorate-General for Communications Networks, Content and Technology is the European Commission department responsible to develop a digital single market to generate smart, sustainable and inclusive growth in Europe.

³Eurostat is the Statistical Office of the European Union.

⁴ECDC: European Centre for Disease Prevention and Control. An agency of the European Union.

Abstract

Countries in Europe took different mobility containment measures to curb the spread of COVID-19. The European Commission asked Mobile Network Operators to share on a voluntarily basis anonymised and aggregate mobile data to improve the quality of modelling and forecasting for the pandemic at EU level. In fact, mobility data at EU scale can help understand the dynamics of the pandemic and possibly limit the impact of future waves. Still, since a reliable and consistent method to measure the evolution of contagion at international level is missing, a systematic analysis of the relationship between human mobility and virus spread has never been conducted. A notable exceptions are France and Italy, for which data on excess deaths, an indirect indicator which is generally considered to be less affected by national and regional assumptions, are available at department and municipality level respectively. Using these information together with anonymised and aggregated mobile data, this study shows that mobility alone can explain up to 92% of the initial spread in these two EU countries, while it has a slow decay effect after lockdown measures, meaning that mobility restrictions seem to have effectively contribute to save lives. It also emerges that internal mobility is more important than mobility across provinces and that the typical lagged positive effect of reduced human mobility on reducing excess deaths is around 14-20 days. An analogous analysis relative to Spain, for which an IgG SARS-Cov-2 antibody screening study at province level is used instead of excess deaths statistics, confirms the findings. The same approach adopted in this study can be easily extended to other European countries, as soon as reliable data on the spreading of the virus at a suitable level of granularity will be available. Looking at past data, relative to the initial phase of the outbreak in EU Member States, this study shows in which extent the spreading of the virus and human mobility are connected. The findings will support policymakers in formulating the best data-driven approaches for coming out of confinement, and mostly in building future scenarios in case of new outbreaks.

Highlights

- this work shows the importance of mobility data in understanding the initial spread of COVID-19;
- it is shown that mobility alone can explain from 52% up to 92% of the excess cumulative deaths;
- it is also shown that mobility correlates with the IgG antibody screening data;
- in the analysed data, internal mobility within a region is more important than mobility across regions;
- the reduction of mobility has a lagged effect of about 14-20 days in flattening the cumulative excess deaths curve;
- further, when containment measures are in place, the impact of mobility on the spread of the virus reduces.

1 Introduction

By means of a letter sent on 8 April 2020, the European Commission asked European Mobile Network Operators (MNOs) to share fully anonymised aggregate mobility data to deliver insights to help fight COVID-19. The aim of the initiative is further specified by the recommendation to support exit strategies policies through mobile data and apps (European Commission, 2020a) with data-driven models that indicate the potential effects of the relaxation of the physical distancing measures (European Commission, 2020b).

Indeed, MNOs can offer a wealth of data on mobility and social interactions. While the value of mobile positioning *personal data* to describe human mobility has been explored (Csáji et al., 2013) and its potential in epidemiology studies demonstrated (Wesolowski et al., 2012, Jia et al., 2020, WU et al., 2020, Kraemer et al., 2020, Wesolowski et al., 2015), the aim of this research is to show that also anonymised and aggregated MNOs data, in compliance with the 'Guidelines on the use of location data and contact tracing tools in the context of the COVID-19 outbreak' by the European Data Protection Board (EDPB, 04/2020), could serve to explain the dynamics of the virus outbreak without the need of disclosing any personal data, nor using tracing applications.

The analysis in this technical report exploits data from four MNOs relatively to three EU Member States (France, Italy and Spain), together with different types of official statistics on the spread of COVID-19 across these countries; the objective is to understand whether or not human mobility matters in the expansion phase of the pandemic.

The report is organised as follows. Section 2 describes the original data provided by the MNOs, namely 'Origin-Destination-Matrices' (ODMs), and the derived mobility indicators used in this study, namely 'Connectivity Matrix'. Section 3 investigates the case of France, taking into account the internal and outbound movements from the Haut-Rhin department (one of the initial hotbeds) to the rest of the country, in order to establish the role of human mobility in explaining the dynamics of the early spread of the virus. The result show that significantly higher excess mortality in other departments can be explained by their connectivity to Haut-Rhin (R^2 up to 0.92). This relationship is valid until mobility is severely reduced due to the national lockdown in force since the 16 March 2020. It also finds out that the curve of cumulative excess deaths in Haut-Rhin correlates almost perfectly (from $R^2 = 0.69$ to $R^2 = 0.98$ after shifting) to the curve of mobility reduction for the same region, showing a lag of 14 days between the two curves. Section 4 focuses on the virus spread in Italy, showing rather similar results to the French case (R^2 up to 0.91 and lagged effect of 14-20 days, depending on the provinces), and confirming that human mobility has an high impact on the virus spread. Since the excess mortality data for Italy is available at municipality level, we have tested the same method also at this higher spatial granularity (the analysis fro France is at Departments level), finding an analogous relationship between mobility and excess deaths in the initial phase of the outbreak; this confirms that the results hold at different geographic scales. Section 5, focuses on Spain, for which instead of the excess death statistics (that are not available at a suitable spatial granularity) we consider the *IgG SARS-Cov-2 antibody screening study* put in place at province level between the end of April and mid May 2020. In this case, only correlation can be tested, and the results show that the number of people found to be positive to IgG test is highly correlated (r up to 75%) with the human mobility derived by the mobile data from the MNOs. It is worth underlining that since, as noticed in (Salje et al., 2020), population immunity appears insufficient to avoid a second wave without lockdown or other containment measures, this work may provide further insights on the effectiveness of containment measures.

Section 6 goes through the main caveats and limitation of the research, pointing out the hypothesis adopted by the proposed analysis.

Finally, Section 7 draws some considerations about the research and highlights the main findings.

2 Mobile Positioning Data and Mobility Indicators

The agreement between the European Commission and the Mobile Network Operators (MNOs) defines the basic characteristics of fully anonymised and aggregate data to be shared with the Commission's Joint Research Centre (JRC⁵). The JRC processes the heterogeneous sets of data from the MNOs and creates a set of mobility indicators and maps at a level suitable to study mobility comparatively across countries; this level is referred to as 'common denominator'.

The following subsections briefly describe the original mobile positioning data from the MNOs and the mobility indicator derived by JRC, respectively.

2.1 Origin-Destination Matrix

Data from MNOs are provided to JRC in the form of Origin-Destination-Matrices (ODMs) (Mamei, 2019, Fekih, 2020). Each cell $[i - j]$ of the ODM shows the overall number of 'movements' (also referred to as 'trips' or 'visits') that have been recorded from the origin geographical reference area i to the destination geographical reference area j over the reference period. In general, an ODM is structured as a table showing:

- reference period (date and, eventually, time);
- area of origin;
- area of destination;
- count of movements.

Despite the fact that the ODMs provided by different MNOs have similar structure, they are often very heterogeneous. Their differences can be due to the methodology applied to count the movements, to the spatial granularity or to the time coverage. Nevertheless, each ODM is consistent over time and relative changes are possible to be estimated. This allows defining common indicators (such as 'mobility indicators' (Santamaria et al., 2020), 'connectivity matrices' and 'mobility functional areas' (Iacus et al., 2020) that can be used, with all their caveats, by JRC in the framework of this joint initiative.

Although the ODM contains only anonymised and aggregate data, in compliance with the EDPB guidelines (EDPB, 04/2020), upon the reception of each ODM, the JRC carries out a 'Reasonability Test'. Both the reasonability test and the processing of the ODM to derive mobility indicators take place within the JRC's *Secure Platform for Epidemiological Analysis and Research* (SPEAR).

This report presents an exploratory analysis for specific outbreaks in France, Italy and Spain, showing and quantifying the relative importance of mobility in the early stages of the virus outbreak in the country. The choice of these three countries is attributed to the availability of ancillary data at high level of granularity in space and time, which are needed for correlation analysis with mobility insights. The mobility within the country is derived using the 'connectivity matrices' that are introduced in the following section.

2.2 Connectivity Matrix

Besides the formulation of 'mobility indicators' (Santamaria et al., 2020) and 'Mobility Functional Areas' (Iacus et al., 2020), an additional example of 'common denominator' to guarantee comparable indicators across countries is a measure of *connectivity* at less granular level, using the common statistical classification of territorial units (NUTS⁶). The NUTS3 level was selected in this framework, where the average population size of administrative units in Member States lie between 150.000 and 800.000.

The connectivity matrix at time t is represented by the Origin Destination Matrix $ODM_{i,j}(t)$ as extracted from normalised and aggregated data received from MNOs. With reference

⁵The Joint Research Centre is the European Commission's science and knowledge service. The JRC employs scientists to carry out research in order to provide independent scientific advice and support to EU policy.

⁶Regulation (EC) 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS)

to Figure 1, considering the NUTS3 area A (rows and columns a_1, a_2 and a_3) and the destination area B (rows and columns b_1, b_2), then the relevant connectivity is calculated as follows:

$$ODM_{A,B}(t) = \sum_{i \in \{a_1, a_2, a_3\}} \sum_{j \in \{b_1, b_2\}} ODM_{i,j}(t) \quad (1)$$

| | | Destination | |
|--------|------|-------------|------|
| | | j=b1 | j=b2 |
| Origin | i=a1 | | |
| | i=a2 | | |
| | i=a3 | | |
| | | | |
| | | | |
| | | | |

Figure 1: Connectivity matrix construction starting from an ODM at higher level of granularity.

The resulting connectivity matrix is obtained by averaging over one week of data. The present study only makes use of outward and internal movements along one row of the matrix and later focuses only on movements internal to departments, i.e. making use only of the diagonal of the connectivity matrix.

3 Human mobility explains the early spread of COVID-19 in France

One of the initial clusters of the COVID-19 outbreak in France is believed to have originated in Haut-Rhin . Since statistics on confirmed cases are by their nature heavily affected by the type and volume of the testing procedures (Bartoszek et al., 2020), which generally varies both across departments and in time as the contagion evolves, we instead have used the total number of (officially confirmed) deaths by departments (see Figure 2 left) over the period 1 March 2020 - 27 April 2020 (INSEE, 2020). France is a special case study thanks to the availability of high resolution deaths counts and our results seem to confirm and complement the analysis in (Salje et al., 2020). In our analysis, we consider the daily cumulative number of excess deaths with respect to the same period of 2019, assuming that most of these excess cases are due to COVID-19 (Wu and McCann, 2020, Giles, 2020, The Economist, 2020). A visual inspection of the data shows that the cumulative sum of excess deaths by department (Figure 2 middle) for the period 1-25 March 2020 is correlated with the connectivity (Figure 2 right) from Haut-Rhin during the week of 23-29 February 2020 (which is supposed to be the start of the pandemic in France). To study the impact of mobility on the spread of COVID-19 in France, we fit a simple statistical model that aims at explaining the excess deaths of Haut-Rhin in terms of human mobility as well as the distance between the department of Haut-Rhin and all other departments. We aim to measure the relative importance of connectivity (or human mobility) compared to the geographical distance for the spread of the virus.

Although mobility can only explain part of the spread of COVID-19, in order to show the relative importance of the mobility component, we run the simple quadratic regression model of equation:

$$y_i^d = const + \alpha_1 \log(\text{mobility}_i) + \alpha_2 [\log(\text{mobility}_i)]^2 + \alpha_3 \text{distance}_i \quad (2)$$

where y_i^d is the number of cumulative death excess for department i from 1 March 2020 till date d ($d=2020-03-01, \dots, 2020-04-27$). The variable distance_i is the geographical distance

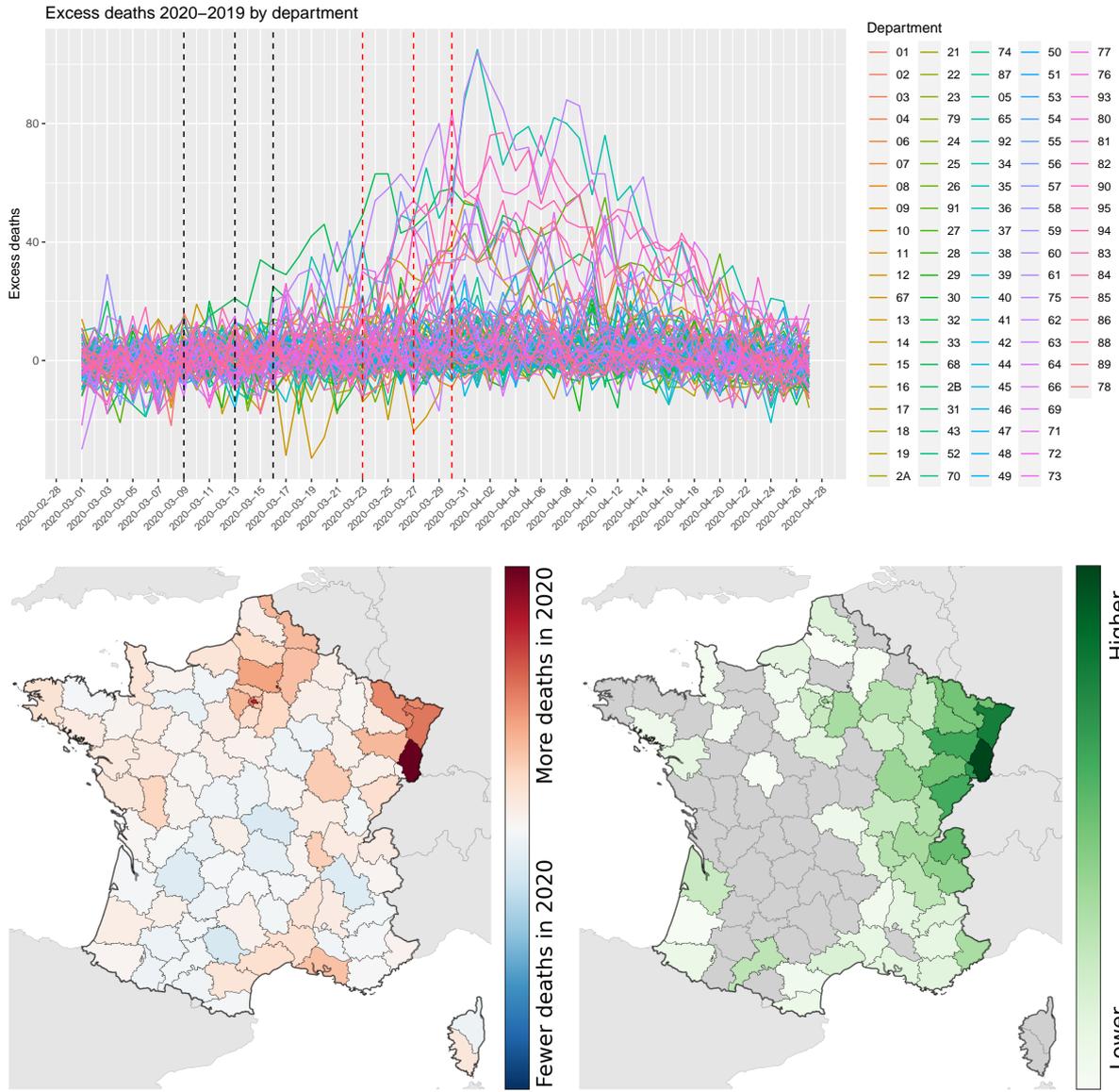


Figure 2: Top panel: The evolution of the excess deaths by department, 1 March - 27 April 2020. Bottom left: Cumulative number of excess deaths 2020-2019, period 1-25 March. Bottom right: connectivity levels from the department of Haut-Rhin (darkest area in the map) during the week of 23-29 February 2020. Grey colour indicates very low connectivity.

between the department Haut-Rhin and the department i and $mobility_i$ is the normalised Mobility Index from Haut-Rhin to the department i for the week of 23-29 February 2020. As we consider log-scaled mobility, we drop the departments for which the outbound mobility from Haut-Rhin is zero, i.e., we only fit this and the other models for which $mobility_i > 0$. The idea behind this data driven approach is to let the data tell us under which conditions and when the mobility data do matter in studying the contagion, and when they do not. The choice of the very simple structure of model Eq. (2) is to capture clearly the impact of mobility, and the scope of the model is not to forecast cases or fatalities but to inform about the dynamics of the COVID-19 due to human mobility. The quadratic form of the model is suggested by the data itself as seen in Figure 4. We also fit two reduced models with *mobility* only Eq. (3) and *distance* only Eq. (4):

$$y_i^d = const + \beta_1 \log(mobility_i) + \beta_2 [\log(mobility_i)]^2 \quad (3)$$

and

$$y_i^d = const + \gamma_1 distance_i. \quad (4)$$

For each model we evaluate the R^2 index as well as the significance of each coefficient. To

further stress the relative importance of the variables in the full model Eq. (2) we consider the evolution of the standardised coefficients in time. Figure 3 shows the goodness of fit of the three competing models in terms of R^2 and the p -value of the estimated coefficients. The vertical lines set on 9 and 13 March represent two large gatherings in the Haut-Rhin region and that on 16 March the first lockdown measure (schools closure). The next three lines are translated by 14 days, which corresponds to the median number of days from the occurrence of the first symptom to death (Lauer et al., 2020). From Figure 3 it is clear that

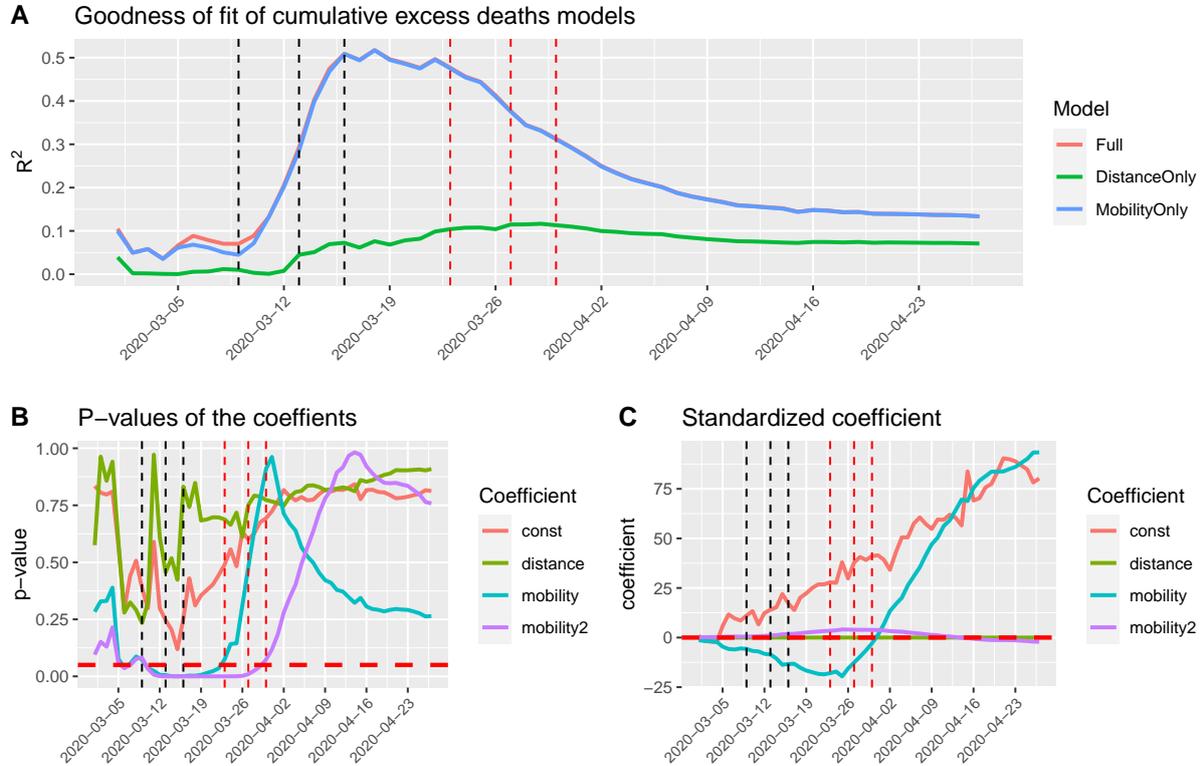


Figure 3: (A) The R^2 is explained mostly by the mobility indicator, and reaches its maximum on 16 March 2020. When the lockdown measures are in force, then the spread is explained by the distance from the Haut-Rhin department, where the spreading of the virus has believed to have started, but also other factors captured by the constant effect. (B) The p -values of the coefficients are presented. The horizontal line represents the 0.05 significance level and (C) the standardised values of the same coefficients to appreciate the relative impact of each on explaining the outcome variable.

the model Eq. (2) dominates the other two but also that, up to the maximum value of the $R^2 = 0.52$ (around 18 March), the model Eq. (3) is almost equally good ($R^2 = 0.51$). Then, on the long run, the distance becomes slightly more important but clearly the rest of the variability is explained by the mix of the three. This figure also supports an argument that the department of Haut-Rhin could have been the initial source of the outbreak of COVID-19 in France. It also demonstrates the positive impact of the lockdown measures on the reduction of excess deaths. Figure 3 also shows the standardised coefficients of the full model Eq. (2) as a function of time. Standardised coefficients are all on the same scale and thus their values can be compared in terms of their impact on the outcome variable. Looking at the plot, then becomes clear that the mobility coefficient is the largest effect up to 27 March, then the excess deaths number is explained by the constant in the model, i.e., by other factors not included in the simple regression analysis.

The explanatory impact of mobility can be improved further if we select those cases for which the mobility index and the cumulative excess deaths are *both* positive. This selection is natural both because what is interesting is the positive excess of deaths and because, due to the process of non-identification and anonymisation of the data, some entries of the original OD matrix are cut to zero by construction. Figure 4 shows the scatterplot of the two dimensions for all the departments and the further selection of data (*cut*). The reference data chosen is 16 March as it is the date for which the $R^2 = 0.92$ of the full model, under the selected sample, is maximal. For the selected data the significance of the coefficients

for the constant and the mobility variable are even improved while for the variable distance is worsened and the standardised coefficients show a similar path to those of Figure 3. Figure 5 also shows that this evidence does not change substantially if we choose a different

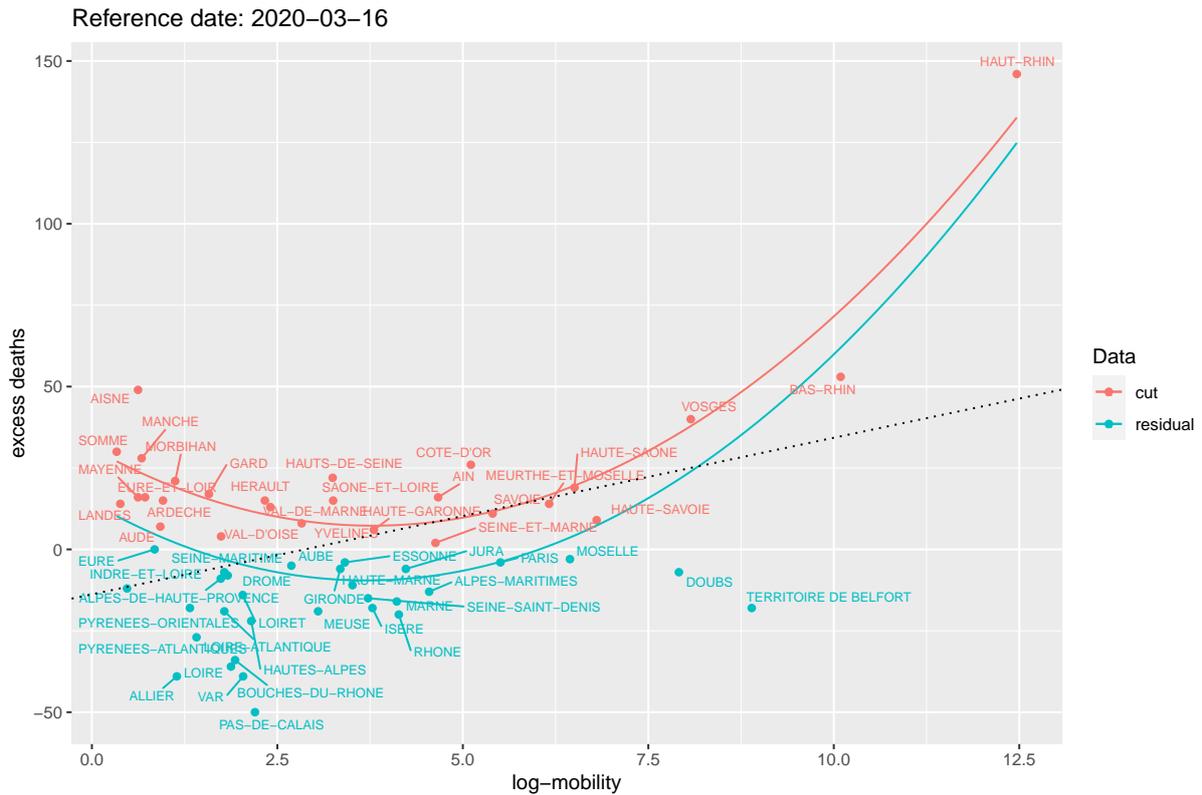


Figure 4: The data for log-mobility and cumulative excess deaths on 16 March 2020. In red the selected subset of positive excess deaths and mobility. The other points are the residual data. Red and light blue together represent the whole data. The two curves correspond to the fitted model Eq. (2) for both data sets. The straight line represent simple linear regression of the two variables in the plot.

mobility week. Indeed, the Figure shows that the R^2 evolution for each mobility week using the model Eq. (3) on the selected data set (*cut*) of positive excess deaths and mobility index present a common pattern, i.e., mobility has no effect before the COVID-19 outbreaks and has a fast decay after the lockdown measures are in place (left and right tails of Figure 5). On the contrary, during the initial emergency phase, mobility is a key element of the spread of the virus and therefore on its outcome (the fatalities). Although, we are not trying to forecast the number of deaths with mobility due to the extreme simplicity of the models considered, we notice that the R^2 path shows an observed time lag between mobility and excess of deaths which we impute to COVID-19: according to (Li et al., 2020) the mean incubation period was 5.2 days (95% confidence interval 4.1 to 7.0) while according to (Wang et al., 2020) the median number of days from the occurrence of the first symptom to death is 14 (range 6 to 41) days. See also (Lauer et al., 2020). We now try to estimate this lag looking at internal mobility.

3.1 Internal mobility matters the most

Most of the correlation captured by the R^2 in the previous analysis is due to the internal mobility within the department and by the connectivity between the Haut-Rhin department and few others (Bas-Rhin, Vosges, Aisne but not, for example, Moselle, see also Figure 4). In fact, the distance alone cannot explain the excess deaths in the chosen simplified Eq. (2) model. Starting from this evidence, which is common to most of the departments considered in the analysis, we now study the impact, if any, of internal mobility alone on the excess deaths of the same department through a time series approach. A similar study using domestic air traffic data at country level can be found in (Zhao et al., 2020). In this work we consider the time series of excess deaths and the one of the reduction of mobility

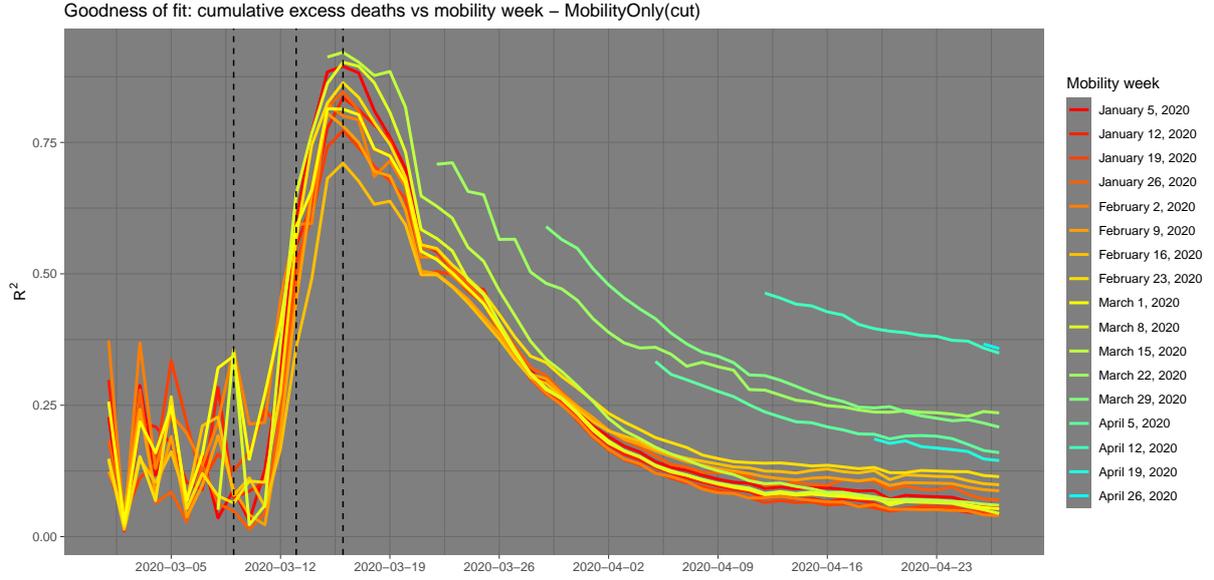


Figure 5: R^2 evolution for each mobility week using the model based on mobility alone on the selected data set of positive excess deaths and mobility index. The peak of correlation is on 16 March 2020 and the mobility week is the one between 15 and 21 March 2020. A similar pattern is shown for all weeks, explaining that mobility has no effect before the COVID-19 outbreaks and has a fast decay after the lockdown measures are in place (left and right tails of the figure). On the contrary, during the initial emergency phase mobility is a key element of the spread of the virus and therefore on fatalities.

(100% = total lockdown, 0% no restriction to mobility) for each department through time. We normalise the reduction of mobility index (namely nmob), to the local maximum within the department to have a $[0,1]$ measure

$$\text{nmob}_{i,j} = 1 - \frac{\text{mobility}_{i,j} - \min_j(\text{mobility}_{i,j})}{\max_j(\text{mobility}_{i,j}) - \min_j(\text{mobility}_{i,j})},$$

for each origin department i and all outbound departments j . We also normalise the cumulative sums of excess deaths in the same way

$$\text{cumdeaths}_i = \frac{\text{cumdeaths}_i - \min_j(\text{cumdeaths}_i)}{\max_j(\text{cumdeaths}_i) - \min_j(\text{cumdeaths}_i)},$$

for each department i .

Figure 6 shows a plot of the two indicators for the Haut-Rhin department. The graphs suggest an evident lag could exist between the two curves. We can estimate the statistical lag in the following way. Let $\theta \in (-\delta, \delta)$ be the time lag between the two nonlinear time series X and Y . Roughly speaking, the idea is to construct a contrast function $U_n(\theta) = \text{Cov}(X_t, Y_{t+\theta})$ which evaluates the Hayashi-Yoshida covariance estimator (Hayashi and Yoshida, 2008, Hayashi and Yoshida, 2005) for the time series X_t and $Y_{t+\theta}$ and then to maximise it as a function of θ . The lead-lag estimator $\hat{\theta}_n$ of θ is defined as (Hoffmann et al., 2013)

$$\hat{\theta}_n = \arg \max_{-\delta < \theta < +\delta} |U_n(\theta)|.$$

When the value of $\hat{\theta}_n$ is positive it means that X_t and $Y_{t+\hat{\theta}_n}$ (or $X_{t-\hat{\theta}_n}$ and Y_t) are strongly correlated, so we say “ X leads Y by an amount of time $\hat{\theta}_n$ ”, so X is the *leader* and Y is the *lagger*. Viceversa for negative $\hat{\theta}_n$. The lead-lag estimator is provided by the *yuima* R package (Iacus and Yoshida, 2018). Applying the lead-lag analysis to our data, we estimate a lag of 14 days which is in line with the median lag time from the onset of symptoms to death (Wang et al., 2020, Lauer et al., 2020). After shifting the normalised mobility curve by 14 days (lagnmob) the cross-correlation between the curve lagnmob and the curve cumdeaths raises from 0.69 to 0.98. A simple linear regression among these two series gives a value of R^2 of 0.96 (from 0.48 before shifting).

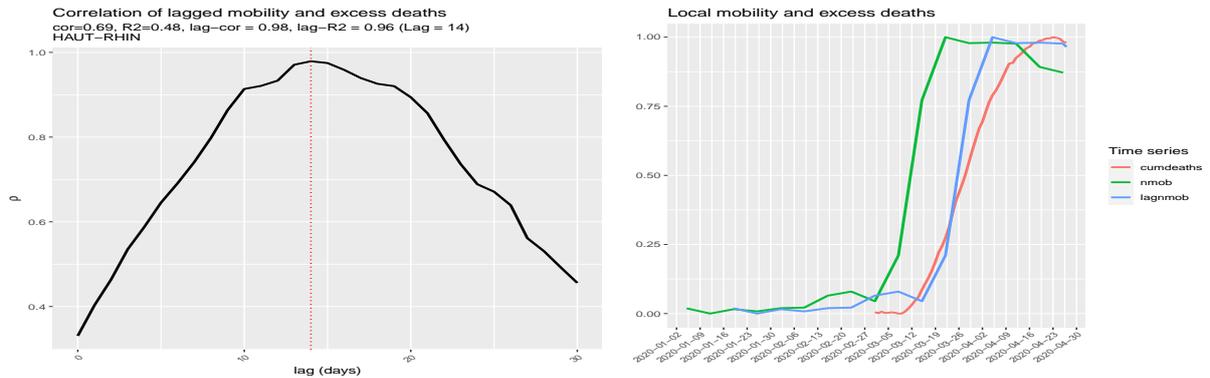


Figure 6: Left panel: the correlation between `cumdeaths` and `nmob` at different lags. Right panel: `cumdeaths`, `nmob` with the additional lagged normalised mobility curve `lagnmob`. The estimated optimal Lag is 14 days.

3.2 Correlation network analysis

After analysing the internal mobility of the Haut-Rhin department, we consider the internal mobility of each department and estimate a lead-lag parameter against the `cumdeaths` curves of all departments. This way of proceedings takes into account possible spin-offs from one department to another but it is not a causal inference analysis. After running the extensive 9216 (96x96) lead-lag analyses, we summarise the results through a network analysis. Therefore, we build a directed graph representing the lead-lag relationships between the *lagger* cumulative deaths curves and the *leader* internal mobility. The nodes in the graph from A_i to B_j are such that B_j is the `cumdeaths` curve for a target department j and A_i is the internal mobility of a department i . Cutting those edges that form loops (internal mobility on the same department) and further selecting the edges $i \rightarrow j$ which presents the maximal lagged-correlation for a given destination j , we obtain a simplified graph. The edges on the graph are weighted according to the R^2 statistics. On this graph, a standard community detection algorithm for directed graphs is run in order to discover similar paths between mobility within a department and impact on other departments. Keeping in mind

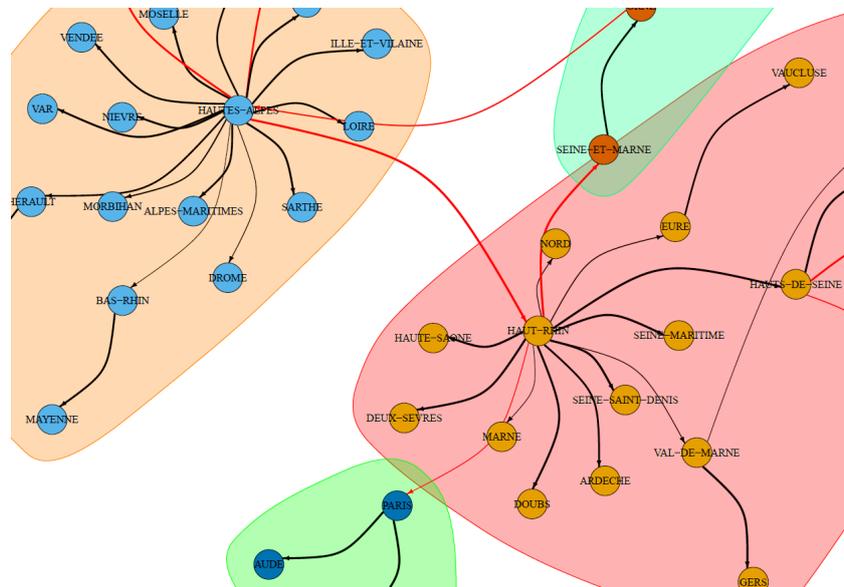


Figure 7: A zoomed section of the graph of lead-lag relationships between internal mobility (origin) and cumulative excess deaths (destination). Full network is shown in Figure 8. The graph is zoomed on Haut-Rhin. This cluster shows that the internal mobility of Haut-Rhin correlates maximally with the cumulative excess deaths of several other departments at different lags. The width of the edges is inversely proportional to the estimated lag: the smaller the lag, the thicker the edge. Red edges connect two different clusters. No causality effect should be read from this graph.

that lead-lag analysis does not involve any causal effect, the clusters that are obtained can

be seen only as similar patterns of lagged cross-correlation between internal mobility and excess deaths in outbound departments. Most of the lag discovered are around 14-20 days. Shorter lag estimates usually correspond to very flat excess cumulative deaths curves. Too large lags (i.e. 30 days) are mostly spurious correlation due to few observations. Figure 7 are a graphical representation of the network graph (the full network is shown in Figure 8.). It is important to stress that no causality effect should be interpreted from this correlation network graph. In fact, notice that the reason why the department of Hautes-Alpes is the origin of the edges of the graph towards many other departments is because the mobility there was reduced much earlier than in other departments and therefore the correlation with the cumulative excess deaths is much anticipated by simple correlative effects. So what can be retained from this analysis? This network says that dynamics of both mobility reduction and cumulative excess deaths have similar patterns regardless of the region. Therefore this result, coupled with previous analysis on internal mobility, can inform policy makers on the expected spread containment in terms of human mobility reduction and the corresponding lag needed to achieve the flattening of the excess deaths curve.

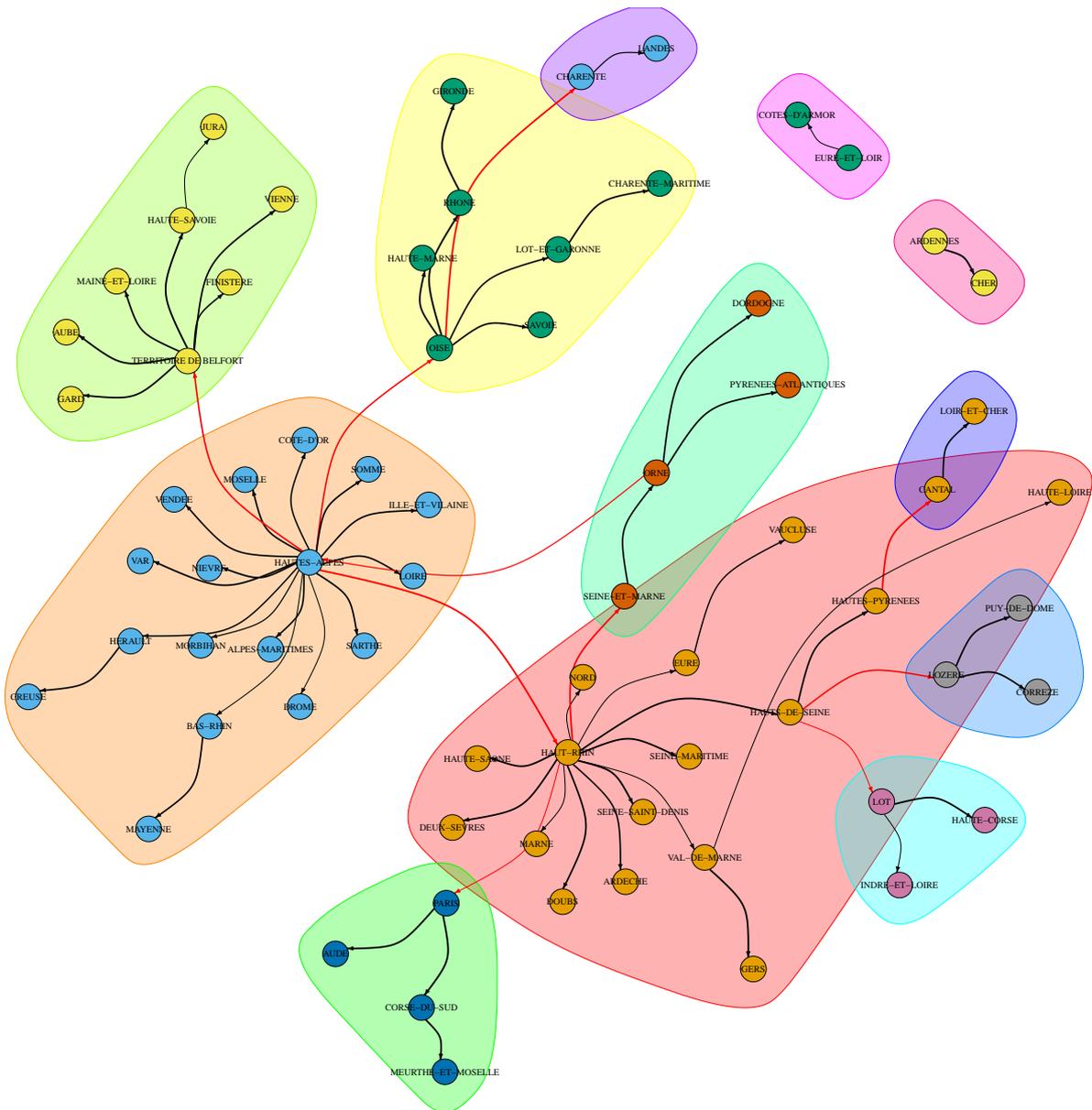


Figure 8: Full graph of lead-lag relationships and cluster of patterns. The width of the edges is inversely proportional to the estimated lag. Red edges connect two different clusters. No causality effect should be read from this graph.

4 Human mobility and CODIV-19 spread in Italy

In this section we have replicated the study of Section 3 for Italy. Data on excess-deaths are publicly available⁷ at (Istat, 2020) at municipality level for 7,270 municipalities (over a total of 7,904, about 92%) for the period 1 January 2020 - 15 May 2020 (see Figure 9). Mobility data for Italy are coming from two different MNOs at different resolution in space (census areas and NUTS3) and time (hourly and daily). The use of both dataset allows to capture short movements (less than the hour, mainly internal mobility or across neighbor provinces) and longer ones (more than one hour) through connectivity matrices even when aggregated at NUTS3 level as considered in our analysis. One of the provinces

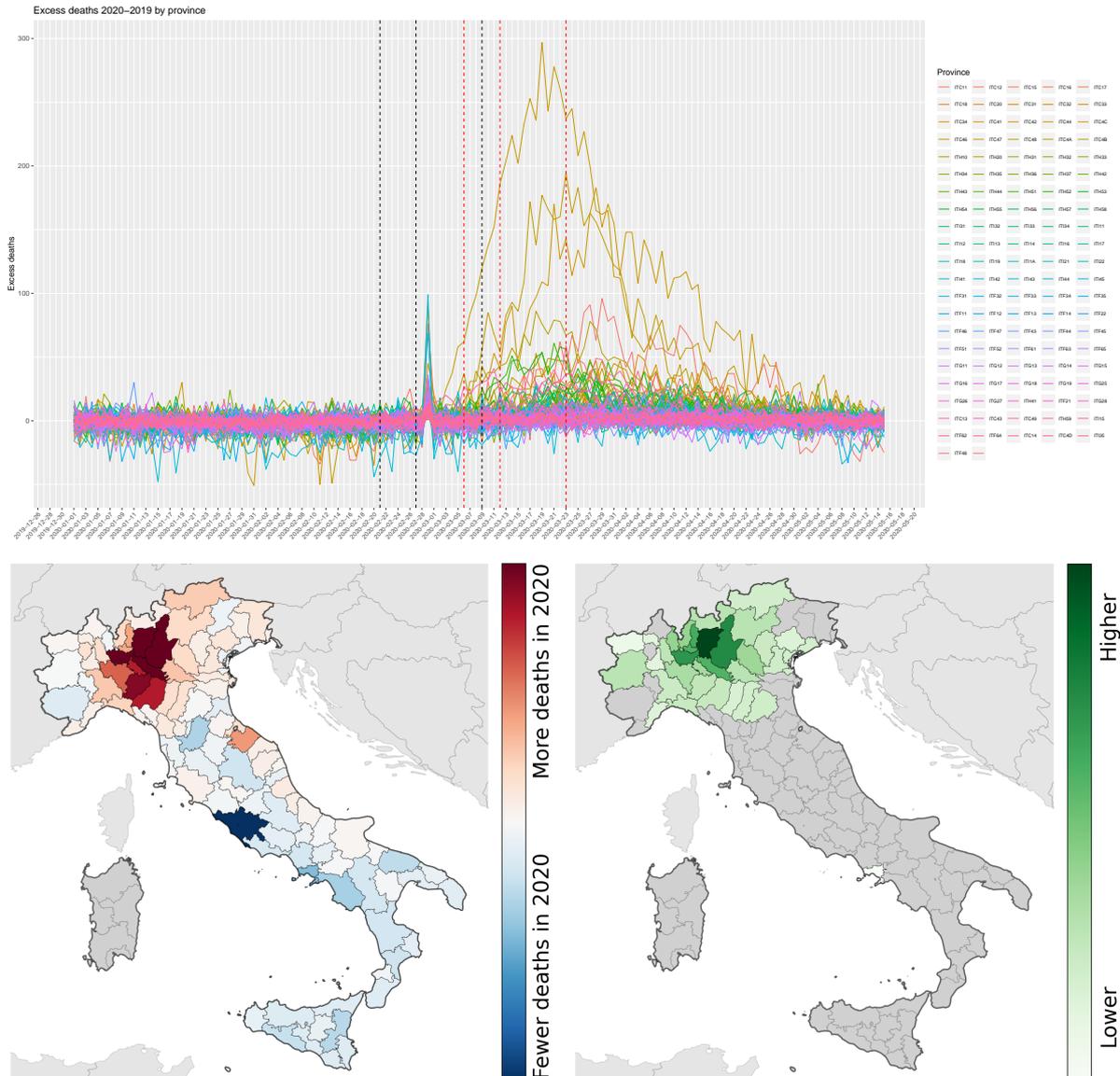


Figure 9: Top panel: The evolution of the excess deaths by province, 1 January - 15 May 2020. Bottom left: Cumulative number of excess deaths 2020-2019, period 1 January-25 March. Bottom right: connectivity levels from the province of Bergamo (darkest area in the map) during the week of 2 March 2020. Grey colour indicates very low connectivity.

strongly hit by the virus initial spread in Italy is Bergamo. We will use this case to show a parallel analysis to that of Section 3. Figure 10 shows the scatterplot for the mobility week of 2 March 2020 against the excess deaths on 1 April 2020. Under the mode model Eq. (3), the $R^2 = 0.91$ for these two set of data. Italy has put in place lock downs for the red zones (very small municipalities) on 21 and 27 February 2020 and the country-wide lock down was enforced on 9 March 2020, so it is interesting to see the pattern of correlation

⁷Data released by Istat on 18 June 2020.

14-20 days after 9 March 2020 as marked with vertical lines in Figure 11. Figure 11 shows

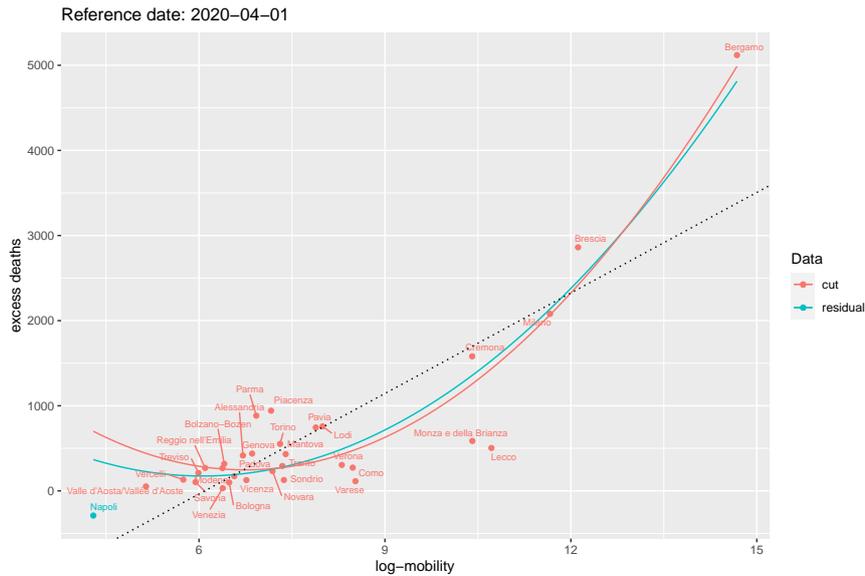


Figure 10: The data for log-mobility on 2 March 2020 and cumulative excess deaths on 1 April 2020. In red the selected subset of positive excess deaths and mobility. The other points are the residual data. Red and light blue together represent the whole data. The two curves correspond to the fitted model Eq. (2) for both data sets. The R^2 for this fitted model is 0.91. The straight line represent simple linear regression of the two variables in the plot. The black vertical lines are on the dates of local lock downs on 21 and 27 February 2020 and country-wise lock down on 9 March 2020. The green lines are the same as the black lines but shifted by 14 days.

that this evidence does not change substantially if we choose a different mobility week. Indeed, the Figure shows that the R^2 evolution for each mobility week using the model Eq. (3) on the selected data set (*cut*) of positive excess deaths and mobility index present a common pattern. In this particular case the R^2 reaches its maximum on 23 March 2020 with the value of 0.95. Now focusing on local mobility via lead-lag analysis, we discover

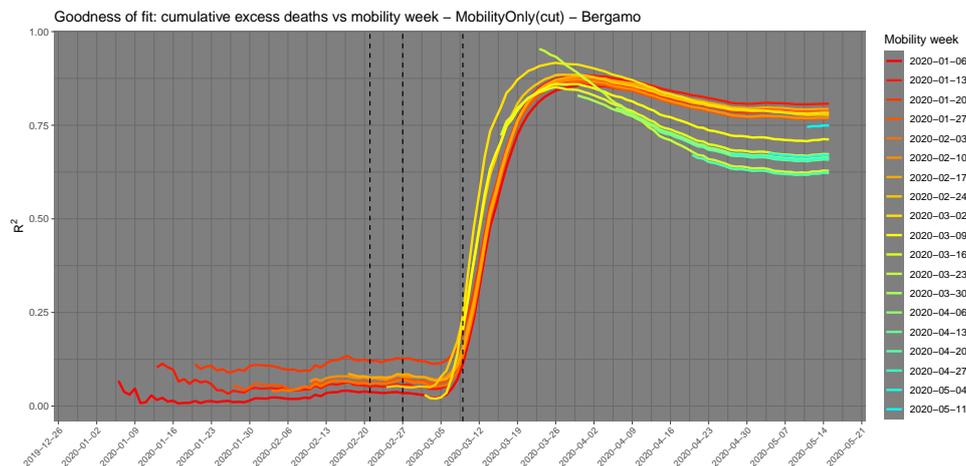


Figure 11: R^2 evolution for each mobility week using the model based on mobility alone on the selected data set of positive excess deaths and mobility index. The peak of correlation is on 23 March 2020 similarly to Figure 5 with $R^2 = 0.95$.

a 18 days lag between mobility reduction and the cumulative curve of excess deaths (see Figure 12). After shifting the normalised mobility curve by 18 days (lag_{mob}) the cross-correlation between the curve lag_{mob} and the curve $cumdeaths$ raises from 0.76 to 0.99. A simple linear regression among these two series gives a value of R^2 of 0.99 (from 0.58 before shifting). The lead-lag analysis shows that, over the 110 Italian provinces, the lagged effect has a mean and median of 20 days, with the first quantile being 16 and third quantile

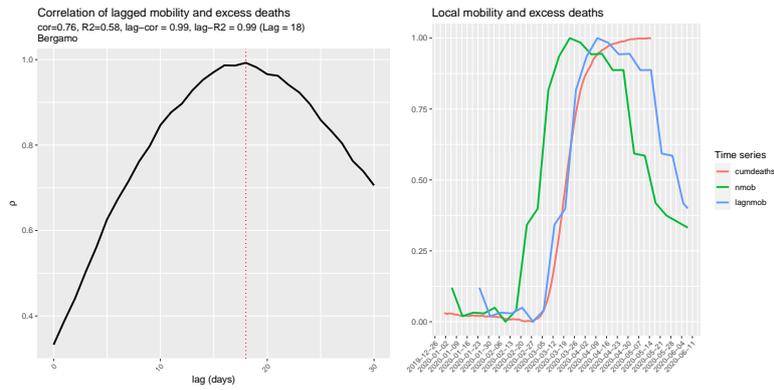


Figure 12: Left panel: the correlation between `cumdeaths` and `nmob` at different lags. Right panel: `cumdeaths`, `nmob` with the additional lagged normalised mobility curve `lagnmob`. The estimated optimal Lag is 18 days.

25 days. Figure 13 presents the correlation network graph for the Italian provinces. As in Section 3 correlation is not causation.

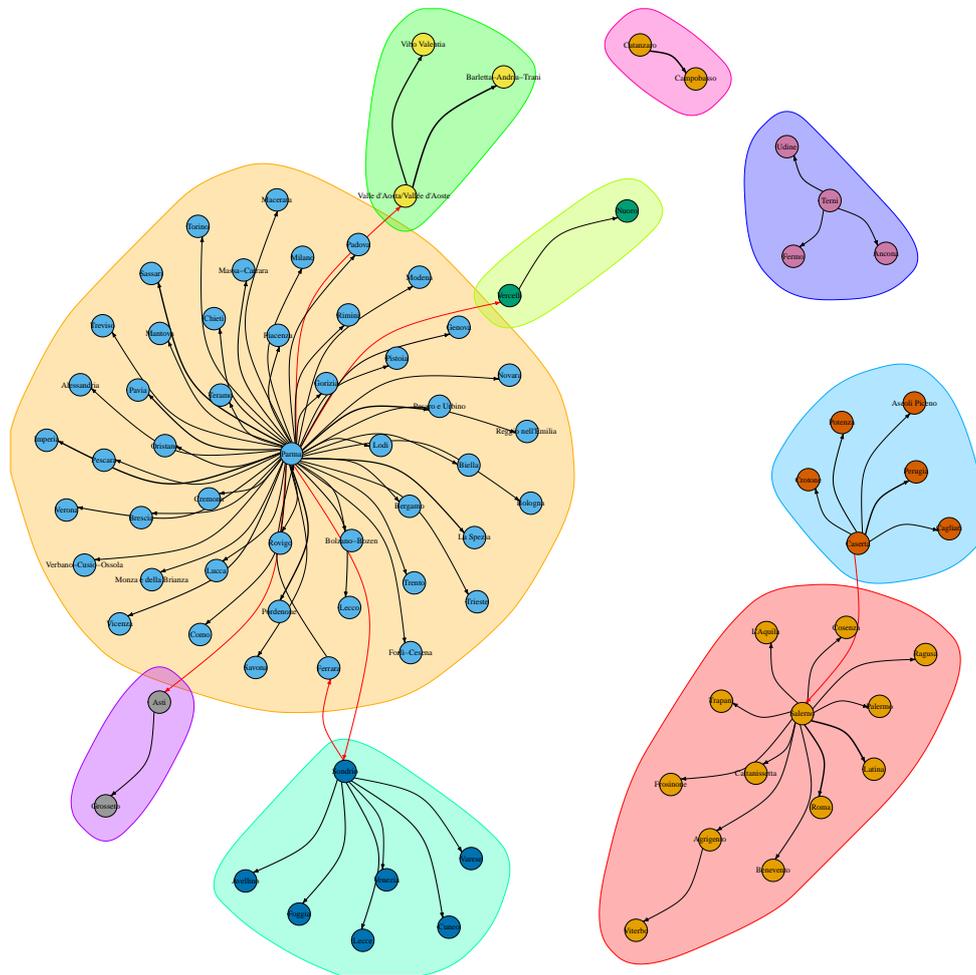


Figure 13: Full graph of lead-lag relationships and cluster of patterns. The width of the edges is inversely proportional to the estimated lag. Red edges connect two different clusters. No causality effect should be read from this graph.

5 Human mobility and IgG antibody testing in Spain

Another interesting case to test the impact of human mobility on the virus spread is Spain. Indeed, despite the fact that the number of cases is known only at region (NUTS2) or country level, this country has conducted a large scale IgG Sars-Cov2 antibody screening on the population at province level in two waves on 27 April 2020 and 11 May 2020. The study, conducted by (ENE, 2020), has recruited 60,983 inland and an additional 3,234 participants in the islands. As running examples, we consider the provinces of Madrid and Barcelona, the two provinces with highest number of cases in Spain, and Badajoz (in Extremadura region) which is an average case. In our study we combine the results of the two weeks of testing

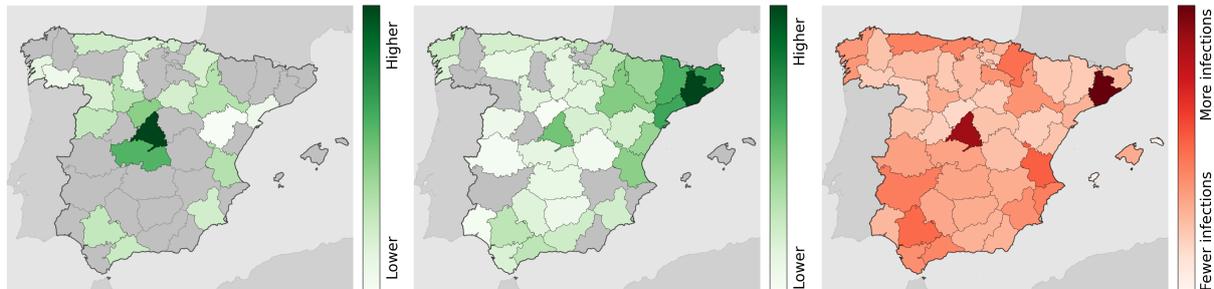


Figure 14: Left and middle panel the mobility from Madrid and Barcelona for the week of 2 March 2020. Right panel the total number of IgG positive cases.

and we compare these data with the mobility data using the same models of the previous sections. Figure 14 shows the mobility from Madrid and Barcelona toward other provinces of Spain in comparison with the number of positive IgG positive cases by province. The combination of the left and middle maps suggests the existence of a relationship between mobility and positive cases that we will now try to quantify through correlation analysis. Figure 15 shows the log-log plot of mobility versus IgG positive cases. The pattern of the relationship between the two variables is similar to what we observed for the excess deaths of France (Figure 4) and Italy (Figure 10).

For Spain though, we don't have time series for the IgG tests, so we cannot run a lead-lag or correlation-network analysis like in Sections 3 and 4, but we can only perform a simple correlation analysis, letting the time of mobility to vary and keeping the IgG data fixed. Figure 16 shows the behaviour of the Spearman (traditional) and Pearson (rank) correlation coefficients between the IgG positive tests and both the mobility and distance between provinces on the linear scale. The evidence here is that mobility is more important, in terms of simple correlation, than the proximity between the provinces up to lockdown which was enforced on Saturday 14 March 2020. The maximal value of the correlation coefficients for Madrid, Barcelona and Badajoz are, respectively, 0.67, 0.56 and 0.56. On the other side, Figure 17 shows the goodness of fit measure R^2 for the Eq. (2) model on the log-log scale. For Madrid and Barcelona $R^2 = 0.57$ while for Badajoz $R^2 = 0.62$. Clearly, the correlations are less strong than those of the excess deaths data because IgG data measure a different aspect of the COVID-19 spread and, begin essentially one point in space, is hard to estimate any variation. Still, the results confirm that human mobility has an impact on the virus spread.

6 Caveats of the study

It is important to remind that the choice to use a simple model of equation Eq. (2) aims at showing the rough effect of mobility on the initial spread of the virus. Clearly, such a model is not intended to and should not be used to produce any other type of result, like, for instance, to forecast the number of deaths. In this work, the excess deaths in France and Italy are treated only as a proxy for the virus spread, as well as IgG testing for Spain is a proxy for the number of people that have been in contact with the virus at a given time. It is also well known that the excess deaths actually due to coronavirus are influenced by a long series of factors such as the age structure of the population and its health condition, the availability of ICU units, the preparedness of the health care system, etc, and that these generally vary across regions. Moreover, part of the excess deaths after lockdown

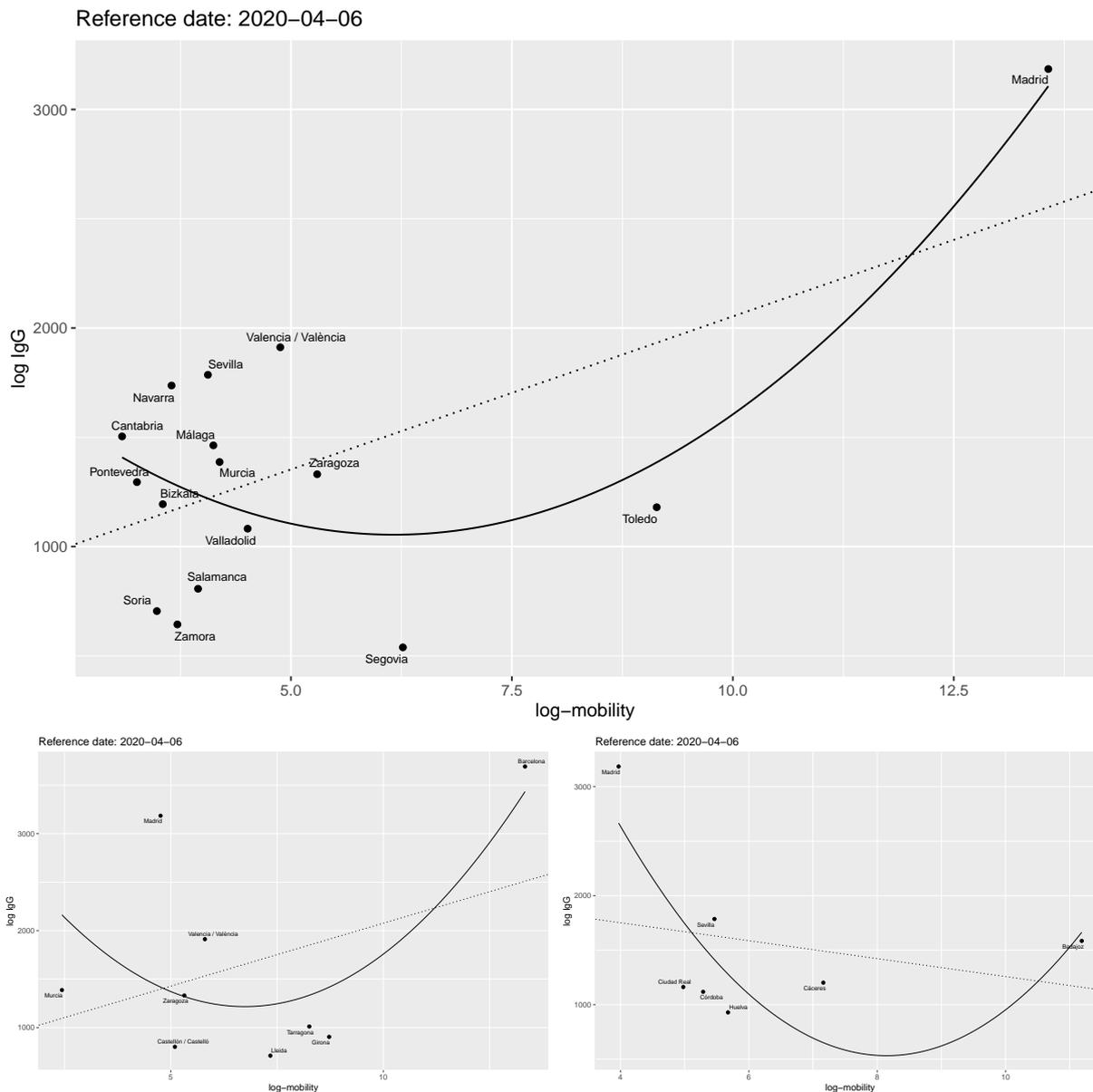


Figure 15: The data for log-mobility and total number of log IgG positive tests for the provinces of Madrid (Top), Barcelona (bottom left) and Badajoz (bottom right). The shape of the relationship is similar to the cases of France (Figure 4) and Italy (Figure 10).

events, may also be deflated by the lower number of deaths due to e.g. road accidents or work-related (many activities have been shut down during national lockdowns). There is also a technical issue due to the construction of the ODMs where all movements below a given threshold are discarded in order to maintain data anonymity.

For obvious reasons (*in primis* for the lack of reliable information) all these aspects have not been accounted for by this study. In other words, despite some of the confounding effects are captured by the constant coefficient of the basic models adopted, clearly there are many other confounders that are not.

Yet, the fact that the results are quite stable and consistent across countries (also adopting different spatial granularities), together with very similar findings published from different research teams relatively to China (Jia et al., 2020), are absolutely encouraging for the evolution of this research.

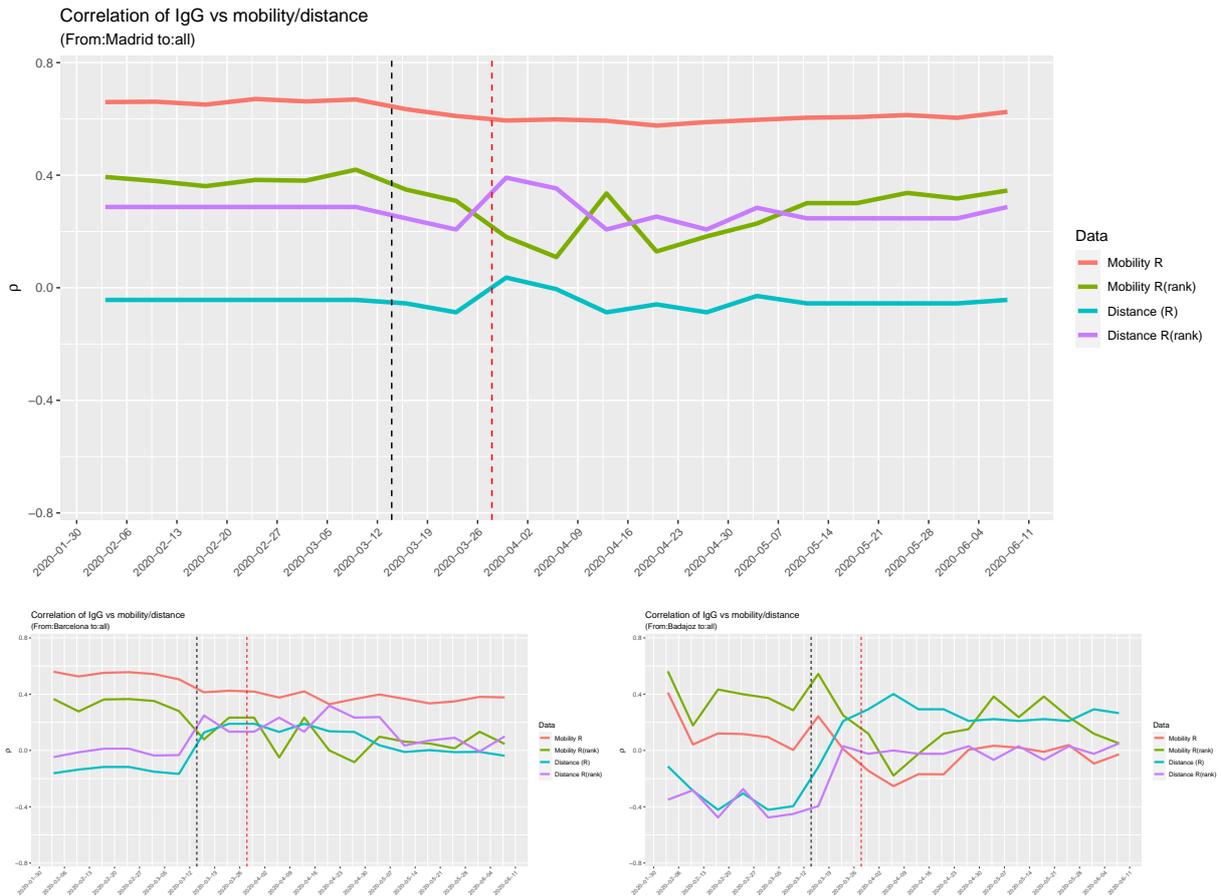


Figure 16: The values of the correlation coefficients for the provinces of Madrid (Top, maximum $|\rho(\text{Mob})| = 0.67$), Barcelona (bottom left, maximum $|\rho(\text{Mob})| = 0.56$) and Badajoz (bottom right, maximum $|\rho(\text{Mob})| = 0.41$), between IgG positive tests and both mobility and distance. The black vertical line correspond to the lock down of 14 March 2020, the read line is just 14 days later.

7 Discussion

This report demonstrates that human mobility, derived by mobile data, is highly correlated with the spread of COVID-19 in the initial phase of the outbreak. In the case study of France, we have found that mobility can explain from 52% up to 92% of the excess deaths reasonably linked to the COVID-19 outbreak. In the case study of Italy, we have found similar results (R^2 up to 0.91 and lagged effect of 14-20 days, depending on the provinces), confirming that human mobility has a high impact on the virus spread, at least before other physical distancing measures are in place. In the case of Spain, we have found that the number of people resulting positive to IgG tests is highly correlated (r up to 75%) with the human mobility.

The above case studies can provide solid evidence to forecasting scenarios for future waves of the virus, in cases where only limited additional protective measures are in place (e.g. wearing masks, physical distancing etc.). This is achieved by exploring the linkage between the geographic distribution of the excess of mortality with respect to 2019 and fully anonymised and aggregated mobile positioning data from European MNOs. Besides predicting dynamics of future COVID-19 outbreaks, connectivity information can be used as basis to plan targeted control measures to curb the spread of the virus. Future data gathered in the context of this European Commission initiative with MNO could enable an analysis at EU regional scale, providing a framework for sharing best practices and data-driven input to inform coherent strategy among EU Member States for de-escalation and recovery.

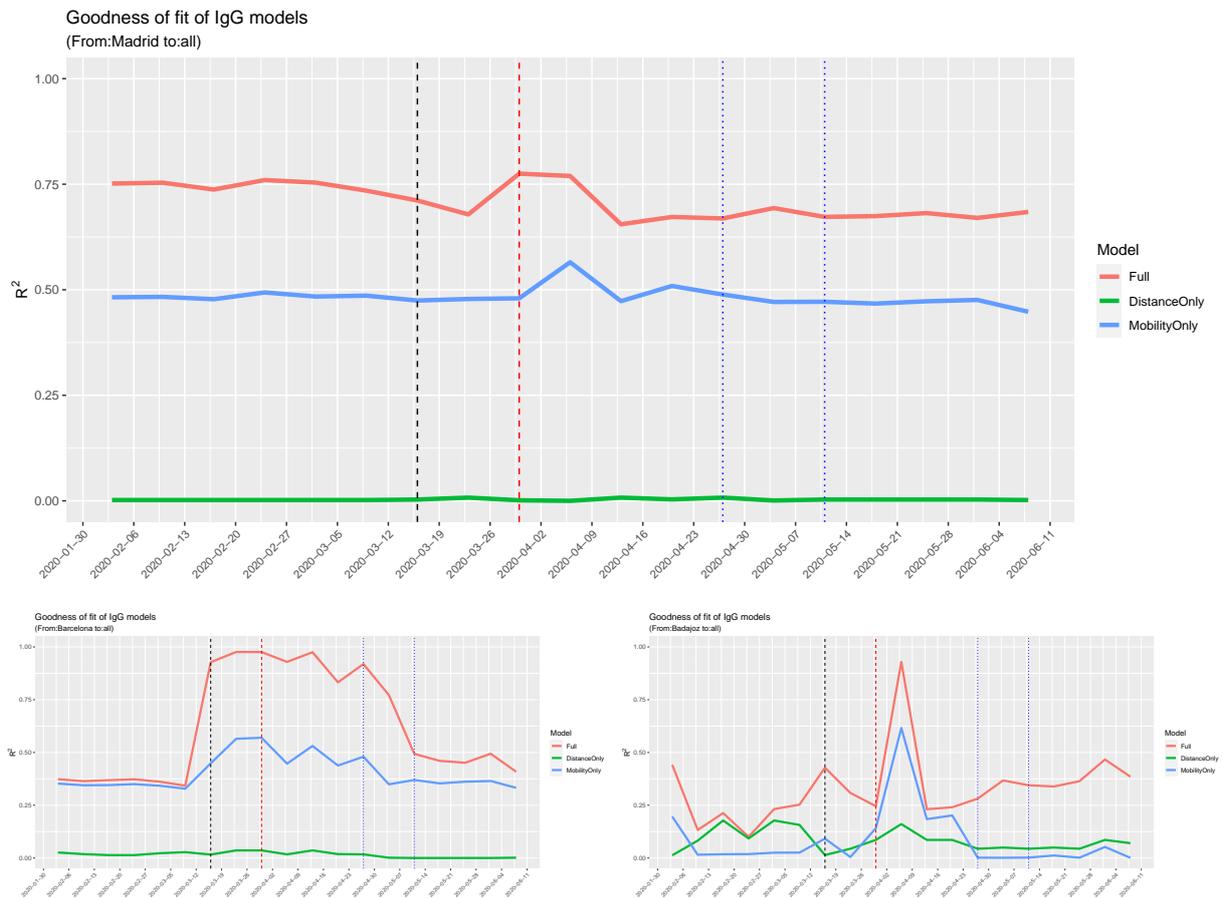


Figure 17: The R^2 values for the provinces of Madrid (Top, maximum $R^2 = 0.57$), Barcelona (bottom left, maximum $R^2 = 0.57$) and Badajoz (bottom right, maximum $R^2 = 0.62$), for the fitted model Eq. (2) between IgG positive tests and both mobility and distance. The black vertical line correspond to the lock down of 14 March 2020, the read line is just 14 days later. The blue lines represent the interval of dates of the IgG screening.

Data and materials availability

Any request of MNO data and derived products should be agreed with each operator, owner of the data.

Data on excess-deaths are publicly available from (INSEE, 2020). Data on excess-deaths are publicly available from (Istat, 2020). Data on the IgG Sars-Cov2 antibody screening in Spain available from (ENE, 2020).

The analysis was entirely conducted with base R with the additional package `yuima`. R code is available upon request to the authors.

References

- Bartoszek, K., Guidotti, E., Iacus, S. M. and Okrój, M., 'Are official confirmed cases and fatalities counts good enough to study the covid-19 pandemic dynamics? a critical assessment through the case of Italy', *Nonlinear Dynamics*, 2020, pp. 1–29. Available at <https://link.springer.com/article/10.1007/s11071-020-05761-w>.
- Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J.-C., Huens, E., Van Dooren, P., Smoreda, Z. and Blondel, V. D., 'Exploring the mobility of mobile phone users', *Physica A: statistical mechanics and its applications*, Vol. 392, No 6, 2013, pp. 1459–1473.
- EDPB, 'Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the covid-19 outbreak', 04/2020. Available at https://edpb.europa.eu/our-work-tools/our-documents/linee-guida/guidelines-042020-use-location-data-and-contact-tracing_en.
- ENE, 'Estudio ene-covid19: Primera ronda estudio nacional de sero-epidemiología de la infección por sars-cov-2 en españa'. 2020. Available at https://www.mschs.gob.es/ciudadanos/ene-covid/docs/ESTUDIO_ENE-COVID19_PRIMERA_RONDA_INFORME_PRELIMINAR.pdf.
- European Commission, 'Commission recommendation (eu) on a common union toolbox for the use of technology and data to combat and exit from the covid-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data, 2020/518'. 2020a. Available at <http://data.europa.eu/eli/reco/2020/518/oj>.
- European Commission, 'The joint european roadmap towards lifting covid-19 containment measures'. 2020b. Available at <https://www.clustercollaboration.eu/news/joint-european-roadmap-towards-lifting-covid-19-containment-measures>.
- Fekih, B. T. S. Z. e. a., M., 'A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of lyon region (france)', *Transportation*, 2020. <https://doi.org/10.1007/s11116-020-10108-w>.
- Giles, C., 'Coronavirus death toll in UK twice as high as official figure'. 2020. Available at <https://www.ft.com/content/67e6a4ee-3d05-43bc-ba03-e239799fa6ab>.
- Hayashi, T. and Yoshida, N., 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli*, Vol. 11, 2005, pp. 359–379.
- Hayashi, T. and Yoshida, N., 'Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes', *Annals of the Institute of Statistical Mathematics*, Vol. 60, 2008, pp. 367–406.
- Hoffmann, M., Rosenbaum, M. and Yoshida, N., 'Estimation of the lead-lag parameter from non-synchronous data', *Bernoulli*, Vol. 19, No 2, 2013, pp. 426–461.
- Iacus, S. and Yoshida, N., 'Simulation and inference for stochastic processes with yuima: a comprehensive r framework for sdes and other stochastic processes', Springer, New York, 2018. ISBN 9783319555690.
- Iacus, S. M., Santamaria, C., Sermi, F., Spyrtatos, S., Tarchi, D. and Vespe, M., 'Mapping mobility functional areas (MFA) using mobile positioning data to inform covid-19 policies, JRC121299', 2020.
- INSEE, 'Nombre de décès quotidiens par département'. 2020. Available at <https://www.insee.fr/fr/information/4470857>.
- Istat, 'Decessi anni 2015-2020: Tavola decessi per 7.270 comuni.' 2020. Available at <https://www.istat.it/it/archivio/240401>.
- Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J. and Christakis, N. A., 'Population flow drives spatio-temporal distribution of COVID-19 in china', *Nature*, 2020, pp. 1–11.
- Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., du Plessis, L., Faria, N. R., Li, R., Hanage, W. P. et al., 'The effect of human mobility and control measures on the covid-19 epidemic in china', *Science*, Vol. 368, No 6490, 2020, pp. 493–497.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G. and Lessler, J., 'The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application', *Ann Intern Med.*, Vol. 172, No 9, 2020, pp. 577–582.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M. and Feng, Z., 'Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia', *New England Journal of Medicine*, Vol. 382, No 13, 2020, pp. 1199–1207.

Mamei, B. N. L. M. e. a., M., 'Evaluating origin–destination matrices obtained from cdr data', *Sensors*, Vol. 19, 2019, p. 1440.

Salje, H., Tran Kiem, C., Lefrancq, N., Courtejoie, N., Bosetti, P., Paireau, J., Andronico, A., Hozé, N., Richet, J., Dubost, C.-L., Le Strat, Y., Lessler, J., Levy-Bruhl, D., Fontanet, A., Opatowski, L., Boelle, P.-Y. and Cauchemez, S., 'Estimating the burden of sars-cov-2 in france', *Science*, 2020. ISSN 0036-8075. . URL <https://science.sciencemag.org/content/early/2020/05/12/science.abc3517>.

Santamaria, C., Sermi, F., Spyrtos, S., Iacus, S. M., Annunziato, A., Tarchi, D. and Vespe, M., 'Measuring the impact of COVID-19 confinement measures on human mobility using mobile positioning data, JRC121298', 2020.

The Economist, 'Tracking covid–19 excess deaths across countries: Official covid–19 death tolls still under-count the true number of fatalities'. 2020. Available at <https://www.economist.com/graphic-detail/2020/04/16/tracking-covid-19-excess-deaths-across-countries>.

Wang, W., Tang, J. and Wei, F., 'Updated understanding of the outbreak of 2019 novel coronavirus (2019-ncov) in wuhan, china', *Journal of medical virology*, Vol. 92, No 4, 2020, pp. 441–447.

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W. and Buckee, C. O., 'Quantifying the impact of human mobility on malaria', *Science*, Vol. 338, No 6104, 2012, pp. 267–270.

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., Engø-Monsen, K. and Buckee, C. O., 'Impact of human mobility on the emergence of dengue epidemics in pakistan', *Proceedings of the National Academy of Sciences*, Vol. 112, No 38, 2015, pp. 11887–11892. ISSN 0027-8424. .

WU, J., Leung, K. and Leung, G., 'Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study', *The Lancet*, Vol. 395, No 10225, 2020, pp. 689–697.

Wu, J. and McCann, A., '25,000 missing deaths: Tracking the true toll of the coronavirus crisis'. 2020. Available at <https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html>.

Zhao, S., Zhuang, Z., Cao, P., Ran, J., Gao, D., Lou, Y., Yang, L., Cai, Y., Wang, W., He, D. and Wang, M. H., 'Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-ncov) cases from wuhan, china in 2020: a correlational analysis', *Journal of Travel Medicine*, Vol. 27, No 2, 02 2020, pp. 1–3. URL <https://doi.org/10.1093/jtm/taaa022>.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

