

# The relationship between human mobility and viral transmissibility during the COVID-19 epidemics in Italy

Paolo Cintia<sup>1</sup>, Luca Pappalardo<sup>2,†</sup>, Salvatore Rinzivillo<sup>2</sup>, Daniele Fadda<sup>2</sup>, Tobia Boschi<sup>3</sup>, Fosca Giannotti<sup>2</sup>, Francesca Chiaromonte<sup>3,4</sup>, Pietro Bonato<sup>6</sup>, Francesco Fabbri<sup>6</sup>, Francesco Penone<sup>6</sup>, Marcello Savarese<sup>6</sup>, Francesco Calabrese<sup>5</sup>, Flavia Riccardo<sup>7</sup>, Antonino Bella<sup>7</sup>, Xanthi Andrianou<sup>7</sup>, Martina Del Manso<sup>7</sup>, Massimo Fabiani<sup>7</sup>, Stefania Bellino<sup>7</sup>, Stefano Boros<sup>7</sup>, Alberto Mateo Urdiales<sup>7</sup>, Maria Fenicia Vescio<sup>7</sup>, Silvio Brusaferro<sup>7</sup>, Giovanni Rezza<sup>7</sup>, Patrizio Pezzotti<sup>7</sup>, Marco Ajelli<sup>8</sup>, Giorgio Guzzetta<sup>8</sup>, Valentina Marziano<sup>8</sup>, Stefano Merler<sup>8</sup>, Piero Poletti<sup>8</sup>, Filippo Trentini<sup>8</sup>, Paolo Vineis<sup>9</sup>, and Dino Pedreschi<sup>1,†</sup>

<sup>1</sup>Department of Computer Science, University of Pisa, Italy

<sup>2</sup>Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Italy

<sup>3</sup>Department of Statistics and Huck Institutes of the Life Sciences, Penn State University, USA

<sup>4</sup>Institute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Italy

<sup>5</sup>Vodafone Analytics

<sup>6</sup>WindTre

<sup>7</sup>Istituto Superiore di Sanità (ISS)

<sup>8</sup>Fondazione Bruno Kessler (FBK)

<sup>9</sup>Imperial College London, School of Public Health

<sup>†</sup>To whom correspondence should be addressed: luca.pappalardo@isti.cnr.it, dino.pedreschi@unipi.it

**In 2020, countries affected by the COVID-19 pandemic implemented various non-pharmaceutical**

**interventions to contrast the spread of the virus and its impact on their healthcare systems and economies. Using Italian data at different geographic scales, we investigate the relationship between human mobility, which subsumes many facets of the population's response to the changing situation, and the spread of COVID-19. Leveraging mobile phone data from February through September 2020, we find a striking relationship between the decrease in mobility flows and the net reproduction number. We find that the time needed to switch off mobility and bring the net reproduction number below the critical threshold of 1 is about one week. Moreover, we observe a strong relationship between the number of days spent above such threshold before the lockdown-induced drop in mobility flows and the total number of infections per 100k inhabitants. Estimating the statistical effect of mobility flows on the net reproduction number over time, we document a 2-week lag positive association, strong in March and April, and weaker but still significant in June.**

**Our study demonstrates the value of big mobility data to monitor the epidemic and inform control interventions during its unfolding.**

## **1 Introduction**

During the COVID-19 pandemic, countries put in place a wide variety of non-pharmaceutical interventions <sup>1–3</sup>, such as closures of activities and travel restrictions, aimed at containing the spread of the virus, mitigating the impact on the healthcare system, and saving lives. Accordingly, citizens adjusted their behavior to the restrictions and the recommendations for personal protection, such as wearing face masks and maintaining social distance. Understanding and monitoring the pan-

demic's evolution in a continuously changing context of policy interventions and behavioral shifts is challenging and yet crucial for policymakers – who need to evaluate the effects of the implemented measures and plan future restrictions or relaxations as the situation unfolds. While different social, economic, demographic, cultural, and psychological variables are at play in the complex and quickly evolving scenario of a pandemic, one physical quantity subsumes many facets of the population's response: *mobility*, i.e. the number of trips, at any scale, that people take to perform their daily activities. Human mobility is a powerful, objective proxy for people's behavior and its adaptation in response to different stimuli during a pandemic, and we have a reliable tool for observing and quantifying human mobility with precision at country scale: mobile phone data<sup>4–8</sup>.

More than a decade of experience in acquisition and analysis of geolocated, spatio-temporal traces left behind by smartphone users has already shown how models based on such data can account for different aspects of our society, also beyond mere mobility<sup>7,9–16</sup>. Mobile phone records provide an unprecedented opportunity to track human displacements<sup>4–7</sup>, allowing one to estimate crucial societal aspects like population density<sup>17–19</sup>, mobility patterns and flows<sup>9,10,20–22</sup>, well-being<sup>13,23–27</sup>, and migrations<sup>20,28–31</sup>. Mobile phone data analytics can be properly designed to preserve privacy and anonymity<sup>7,32–35</sup>, and is advocated also as a resource for supporting public health actions across the phases of the COVID-19 pandemic<sup>36,37</sup>. Motivated by the potential of mobile phone data in capturing the geographical spread of epidemics<sup>38–41</sup>, researchers and governments have started to collaborate with mobile network operators to estimate the effectiveness of control measures in several countries<sup>42–55</sup>.

In this paper, we investigate in depth the relationship between human mobility and the spread of COVID-19 at different geographic scales, leveraging nation-wide Italian mobile phone data provided by two major telecom operators, Vodafone and WindTre, spanning 9 months since the start the outbreak of the virus.<sup>a</sup> We compare the evolution of daily mobility flows and the evolution of viral transmissibility, measured by the net reproduction number (mean number of secondary infections generated by one primary infector) from January through September 2020 – a period in which several control interventions and human behavioral adaptations took place across the country. As a key result, we find a striking relationship between the decrease in mobility flows and the decrease in net reproduction number in all Italian regions between March 11th and March 16th, when the country entered the national lockdown. We can quantify the time needed to switch off mobility and the time required to bring the net reproduction number below the critical threshold of 1 (one week).

We also find a strong relationship between the number of days spent above the epidemic threshold before the lockdown-induced drop in mobility flows and the total number of confirmed SARS-CoV-2 infections per 100k inhabitants. While many other factors may have played a role in containing the contagion, including other non-pharmaceutical interventions, this observation provides strong evidence supporting the effectiveness of mobility restrictions. Finally, we use Functional Data Analysis tools to estimate the statistical effect of mobility flows on the net reproduction number over time. We document a 2-week lag positive association that was very strong in March and April and weaker but still significant in June.

---

<sup>a</sup>From February till September 2020. We are currently extending the analysis for the remaining months of 2020.

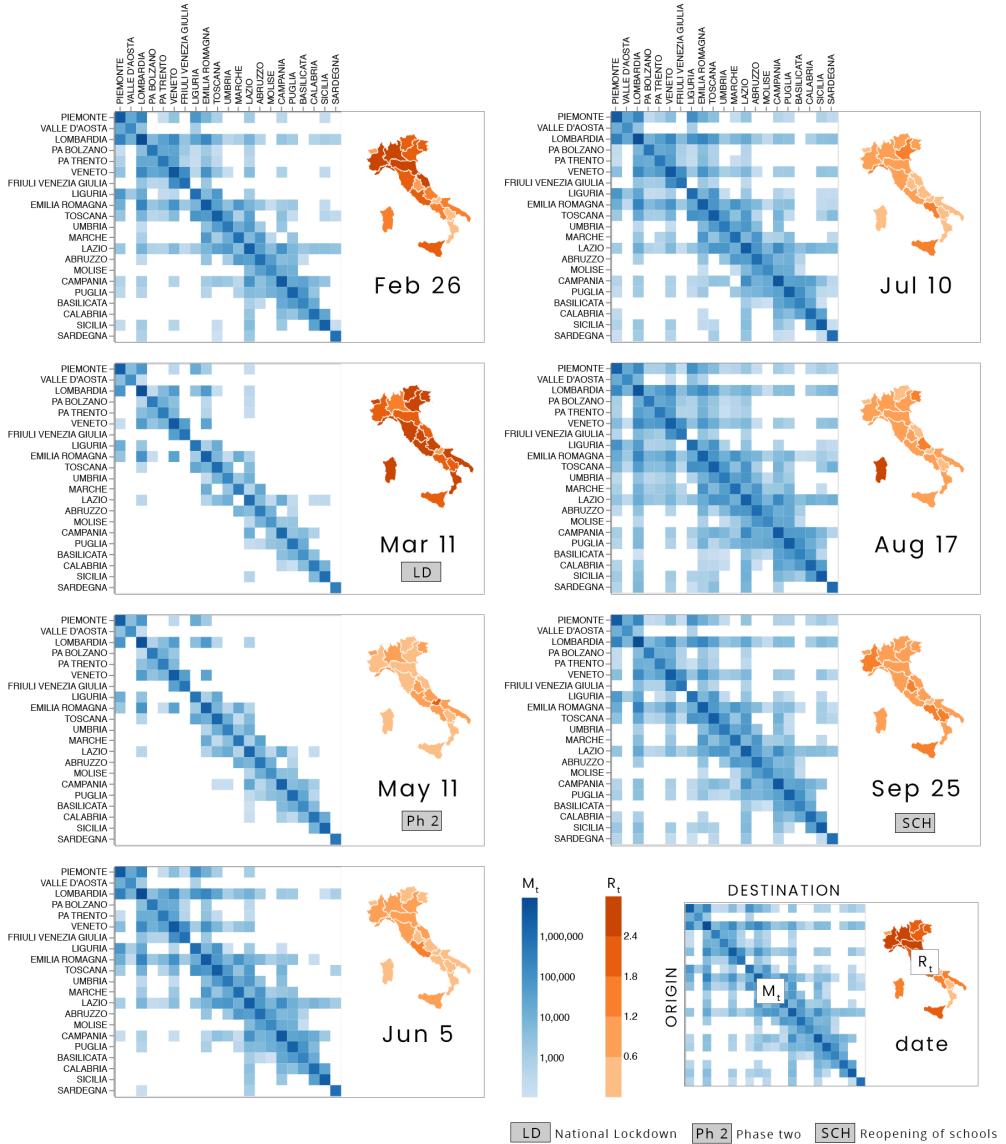
Understanding the relationship between human mobility patterns and the spread of COVID-19 is crucial, in particular, for restarting social and economic activities, which were limited or put in stand-by during national and regional lockdowns, as well as to monitor the risk of a resurgence during lockdown exits. Our study, developed within the context of the Italian data-driven task force to support policy making during the COVID-19 epidemics, demonstrates the value of "big" mobility data for monitoring the epidemic and inform control interventions during its unfolding.

## 2 Data sets

**Mobility data.** We use data derived from the records the two largest mobile phone operators in Italy: WindTre and Vodafone, which cover together approximately two thirds of the Italian mobile phone users. The basic geographical unit is the phone cell, defined as the area covered by a single antenna, i.e., the device that captures mobile radio signals and keeps the user connected with the phone network. Multiple antennas are usually mounted on the same tower, each covering a different direction. The tower's geographic position and the antenna's direction allow one to infer the extension of the corresponding phone cell. A user's position is approximated by the antenna serving the connection, whose extension is relatively small in urban contexts (in the order of 100m  $\times$  100m) and larger in rural areas (in the order of 1km  $\times$  1km or more). A user's mobility is reconstructed as the sequence of towers their phone is connected to in time. A connection (event) is created every time the user calls or makes an upload/download operation from the Internet. We gather such data for the period from February 01st, 2020, to September 30th, 2020.

We reconstruct daily Origin-Destination matrices (ODs)<sup>22,56,57</sup> in which the flow between two municipalities  $i$  and  $j$  is the number of individuals that moved from  $i$  to  $j$  in a given day. To reduce noise, we consider only movements for which the individual remains in the destination municipality for at least 30 minutes. For privacy-preserving purposes, we remove from the OD matrices flows with less than 15 trips between two municipalities. To match the available COVID-19 data, we aggregate the municipality-to-municipality ODs into province-to-province and region-to-region ODs, in which each node represents an Italian province or region. In particular, for each day, we compute both the in-flows, which capture the total number of people moving to a province/region from any other province/region, and the self-flows, which capture trips between municipalities of the same province/region. We then sum in-flows and self-flows to form a daily time series  $M_t$  for each province/region. Figure 1 shows the structure of the region-to-region OD matrices for different days in the period of observation. As expected, we found no statistically significant differences between the OD matrices from the two mobile phone operators, to the purpose of the analyses presented in this paper. This observation brings robustness to the results, and allows use to use any of the two sources.

**Epidemiological data.** For each Italian province/region, we consider the daily time series  $R_t$  of the net reproduction number, which represents the mean number of secondary infections generated by one primary infector. When  $R_t$  decreases below the epidemic threshold of 1, the number of new infections begins to decline.  $R_t$  was estimated from the daily time series of new cases by date of symptom onset. Case-based surveillance data used for estimating  $R_t$  were collected by regional health authorities and collated by the Istituto Superiore di Sanità (ISS) using a secure



**Figure 1:** Evolution of mobility flows between origin and destination regions, and regional  $R_t$  levels, for a selection of seven days in the period February 2020 - September 2020. The adjacency matrix on the left of each panel depicts mobility flows, with a color intensity proportional to the number of trips between origin and destination. Both flows and heterogeneity in destinations decreased during lockdown or phase 2 (e.g., March 11, May 11), with a tendency for flows to persist within neighboring geographical clusters. The choropleth maps of Italy on the right of each panel show the intensity of  $R_t$  for each region.

online platform, according to a progressively harmonized track-record. This data includes, among other information, the place of residence, the date of symptom onset, and the date of first hospital admission for laboratory-confirmed COVID-19 cases<sup>58</sup>. The distribution of  $R_t$  was estimated with a well-established statistical method<sup>59–61</sup> based on the knowledge of the distribution of the generation time and on the time series of cases. In particular, the posterior distribution of  $R_t$  for any time  $t$  was estimated by applying Metropolis-Hastings MCMC sampling to the likelihood function

$$L = \prod_{t=1}^T P\left(C(t); R_t \sum_{s=1}^T \phi(s) C(t-s)\right) \quad (1)$$

where  $P(x; \lambda)$  is the probability mass function of a Poisson distribution (i.e., the probability of observing  $x$  events if these events occur with rate  $\lambda$ );  $C(t)$  is the daily number of new cases having symptoms onset at time  $t$ ;  $R_t$  is the net reproduction number at time  $t$  (the quantity to be estimated); and  $\phi(s)$  is the probability distribution density of the generation time evaluated at time  $s$ .

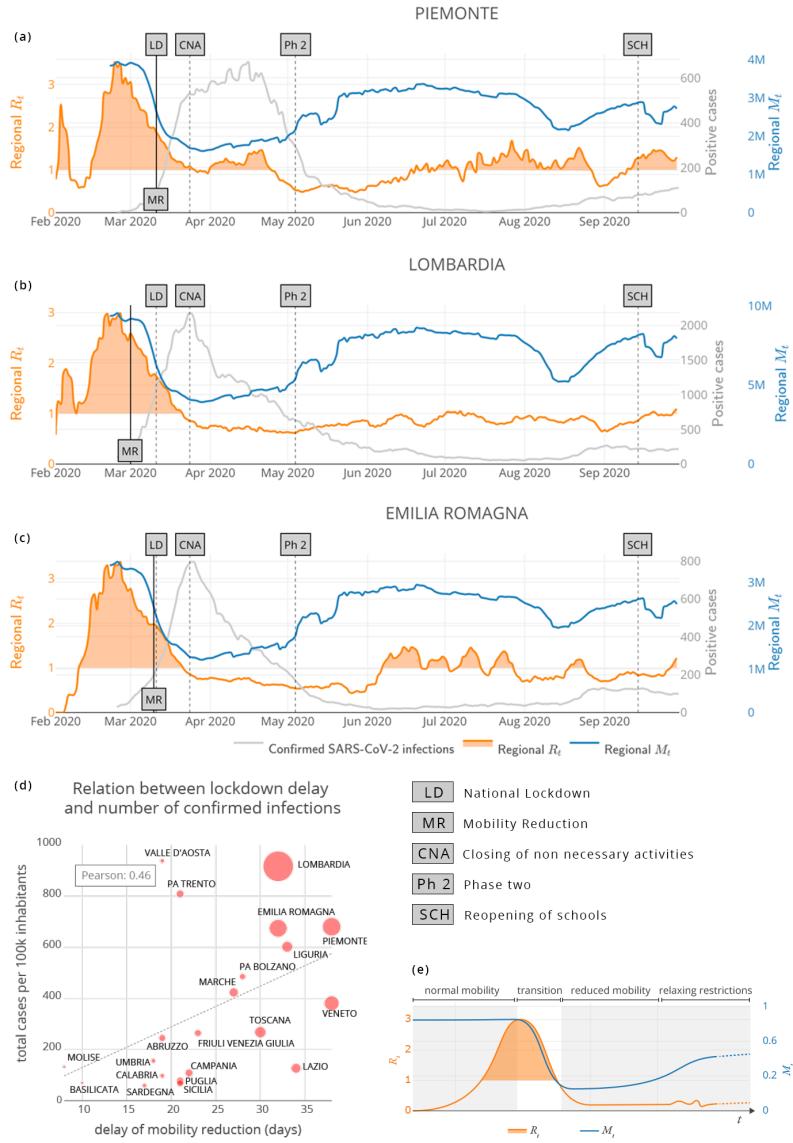
As a proxy for the distribution of the generation time, we used the distribution of the serial interval, estimated from the analysis of contact tracing data in Lombardy<sup>58</sup>, i.e., a gamma function with shape 1.87 and rate 0.28, having a mean of 6.6 days. This estimate is within the range of other available estimates for SARS-CoV-2 infections, i.e., between 4 and 7.5 days<sup>62–64</sup>. The number of reported SARS-COV-2 infections is provided by Protezione Civile, the Italian public institution in charge of monitoring the COVID-19 emergency. They collect data from every Italian administrative region and make them available on a public repository. For each province/region, we focus on the number of new positive cases per day. Specifically, given a day  $g$ , we average the values over the four days before and the four days after  $g$ . Figure 1 shows the value of  $R_t$  for

Italian regions in different days of the observation period.

### 3 Relationship between $M_t$ and $R_t$

Figure 2 shows  $M_t$  (blue curves),  $R_t$  (orange curves) and the number of positive cases (grey curves) for three Italian regions (similar plots for all other regions are provided in Supplementary Figure S1). In all regions,  $M_t$  decreases sharply soon after the first national lockdown (March 11th) and stabilizes on a new, reduced level after about one week. Subsequent restriction ordinances, such as the closing of non-essential economic activities on March 17th, have a minor impact on the reduction of  $M_t$ . In almost all regions,  $M_t$  increases at the start of “phase 2” (partial lifting of the lockdown) on May 4th. This behaviour is particularly pronounced for Emilia-Romagna (Figure 2c), Toscana, Puglia and Lazio (Supplementary Figure S1 j,m,q). Interestingly, we find a slight increase of  $M_t$  approaching May 4th. This may be due to a progressive, although slight, relaxation of compliance with the mobility limitations imposed by the lockdown. The case of Molise is different and particularly compelling: it is indeed the only region for which  $M_t$  decreases after May 4th (Supplementary Figure S1 o). This may be due to news media coverage of a funeral on April 30th, attended by a large number of people, which resulted in a large local outbreak. The news may have induced parts of the population of Molise to self-restrict movements during the following days.  $M_t$  returns to the original level between May and June, around two weeks after the starting of phase 2.

Although the date when  $R_t$  surpasses 1 for the first time varies from region to region, for most



**Figure 2:** Evolution of  $M_t$  (blue),  $R_t$  (orange, 7-day moving average) and number of confirmed infections (grey; 9-day moving average) between Feb - Sep 2020 for Piemonte (a), Lombardia (b), Emilia-Romagna (c). Orange-shaded areas mark periods where  $R_t > 1$ . Vertical dashed lines indicate: beginning of national lockdown (LD, March 9th, 2020), closing of non-essential economic activities (CNA, March 23th, 2020), partial restart of economic activities and within-region mobility (“Ph 2”, May 4th, 2020), and beginning of the school year (SCH). In panel (d), regions are represented in terms of the number of days between the first occurrence of  $R_t > 1$  and the beginning of the national lockdown (horizontal axis), and the cumulative incidence of confirmed SARS-CoV-2 infections per 100k inhabitants as of May 15th, 2020 (vertical axis). The size of the circles is proportional to the total number of positive cases in the period. A linear least square fit (dotted line) indicates a positive relationship. Panel (e) schematizes the epi-mob pattern: during the first week of lockdown, the two curves describing mobility flows and net reproduction number gracefully overlap. At the end of this week, the country reaches a new mobility regime at  $\approx 40\%$  of the pre-lockdown level.

of the regions  $R_t$  decreases concurrently with  $M_t$  due to the beginning of the national lockdown (see Figure 2, Supplementary Figure S2), highlighting the importance of governmental interventions to reduce mobility. Note that, for most of the regions,  $R_t$  reaches values lower than 1 starting around March 16th, when  $M_t$  stabilizes on the new, reduced level until May (Supplementary Figure S1).  $M_t$  increases between May and June. Yet, in this period,  $R_t$  continues to decrease. This may be due to human mobility not coming back to the usual volume and connectivity, as well as to other ordinances by local and national governments related to the wearing of masks and gloves in public areas, social distancing and ban on gatherings – and possibly to other factors to be further investigated. While  $M_t$  remains stable until mid August, in many regions  $R_t$  starts slightly increasing again from June onward, though with different patterns in different regions. This result indicates that releasing mobility restrictions for a long enough time boosts the circulation of the virus, highlighting *a fortiori* the relationships between mobility and viral transmissibility.

We go one step further in our analysis of the relationship between mobility flows and contagion by computing two quantities for each region: (*i*) the delay in mobility reduction, i.e., the number of days in which  $R_t > 1$  before the regional  $M_t$  decreases by at least 20% w.r.t. the usual (pre-epidemics) weekly mobility, observed over the first two weeks of February; and (*ii*) the total number of reported SARS-CoV-2 infections per 100k inhabitants in the region (as of May 15th, 2020, when national mobility restrictions were released).

The date of mobility reduction below 20% for each region is indicated as the MR black vertical line in Figures 2(a), (b) and (c). As shown in Figure 2(d), the two quantities are positively

and significantly correlated (Pearson coefficient= 0.46,  $p < 0.05$ ,  $r^2 = 0.21$ ) – suggesting that larger delays could have induced heavier spreading of the virus. This is strong evidence that timely lockdowns are instrumental for better containment of the contagion.

The delay in mobility reduction in Lombardy was 32 days, leading to the highest number of positive cases per inhabitant in Italy. Similarly, for other regions severely affected by the virus, such as Liguria, Emilia-Romagna and Piemonte, the delays in mobility reduction were 32 and 38 days, respectively. The north-central and north-western regions all lie above the dashed regression line in the top right portion of the scatter plot in Figure 2(d). Regions below this line in the bottom right portion of the plot, such as Veneto, Lazio and Tuscany, were more effective in containing the contagion, despite delays in mobility reduction of 30 days or more. This fact may be explained by several factors, including the effectiveness of the epidemic surveillance, the intensity of the testing and tracing strategy adopted, the capacity of outbreak containment, and also the absolute number of cases when  $R_t$  jumps above 1. On the other hand, southern regions all had smaller delays in mobility reduction – in the range of 10 (Molise, Basilicata) to 20 (Campania, Puglia, Sicily, Calabria) days. This appears to have helped in containing the spread of the virus; all southern regions (with the only exception of Molise) are below the regression line in the bottom left portion of the plot, with low numbers of infections per 100K inhabitants. In summary, the regions that appear to have benefited the most from the lockdown are those where the reduction in mobility occurred in a more timely fashion.

## 4 Lags and Function-on-function regression

Another way to characterize the relationship between  $M_t$  and  $R_t$  is to employ statistical tools from Functional Data Analysis (FDA). These allow us to efficiently exploit the longitudinal information at our disposal, evaluating lags and regressing transmissibility onto mobility. First, we smooth the discrete observations of  $R_t$  and  $M_t$  using B-splines basis functions of order 4<sup>65</sup>. To remove daily fluctuations and weekly trends, we use 32 basis functions – one per week – and add a roughness penalty on the second derivative of the curves. We select the smoothing parameter minimizing the average generalized cross-validation error across all the curves<sup>66</sup>. Panels (a) and (b) in Supplementary Figure S2 show smoothed  $R_t$  and  $M_t$  curves for the twenty Italian regions, respectively. Based on the averages of these two sets of curves (black solid lines), the delay between the  $M_t$  peak (late February) and the  $R_t$  peak (early March) is approximately 2 weeks – 13 days. A similar lag is supported considering the projections of the curves on their first functional covariance component (FCC)<sup>67</sup>. The first FCC identifies a reduced functional space which captures the most important mode of covariation between  $R_t$  and  $M_t$ , and allows us to further de-noise the data. Panels (c) and (d) in Supplementary Figure S2 show the FCC projections together with their averages (again black solid lines). The peaks are again 13 days apart.

Next, we seek an optimal alignment (or *registration*) of the curves which, separating horizontal and vertical variation, increases the statistical power of the analysis<sup>65</sup>. Specifically, for each region, we estimate the horizontal shift (capped at a maximum of 20 days) that minimizes the L2 distance  $\int (R(t) - \tilde{R}(t))^2 dt$  between the  $R_t$  curve and its FCC projection  $\tilde{R}(t)$  (FCC projections

are an ideal alignment target here since they do not exhibit any horizontal variation; see Figure S2).

We apply the estimated shifts to the  $R_t$  as well as the  $M_t$  curve in each region, preserving their time consistency. The shifted  $R_t$  and  $M_t$  curves restricted to the intersection of their new time domains are displayed in Figure 3 panels (a) and (b), respectively. The estimated shifts are shown in Supplementary Figure S2 panel (e). One can note how the registration procedure translates the epidemic curves aligning the  $R_t$  peaks.

Finally, we apply function-on-function regression<sup>68</sup> to the registered curves to estimate the statistical effect of  $M_t$  on  $R_t$ . In symbols, we fit the model:

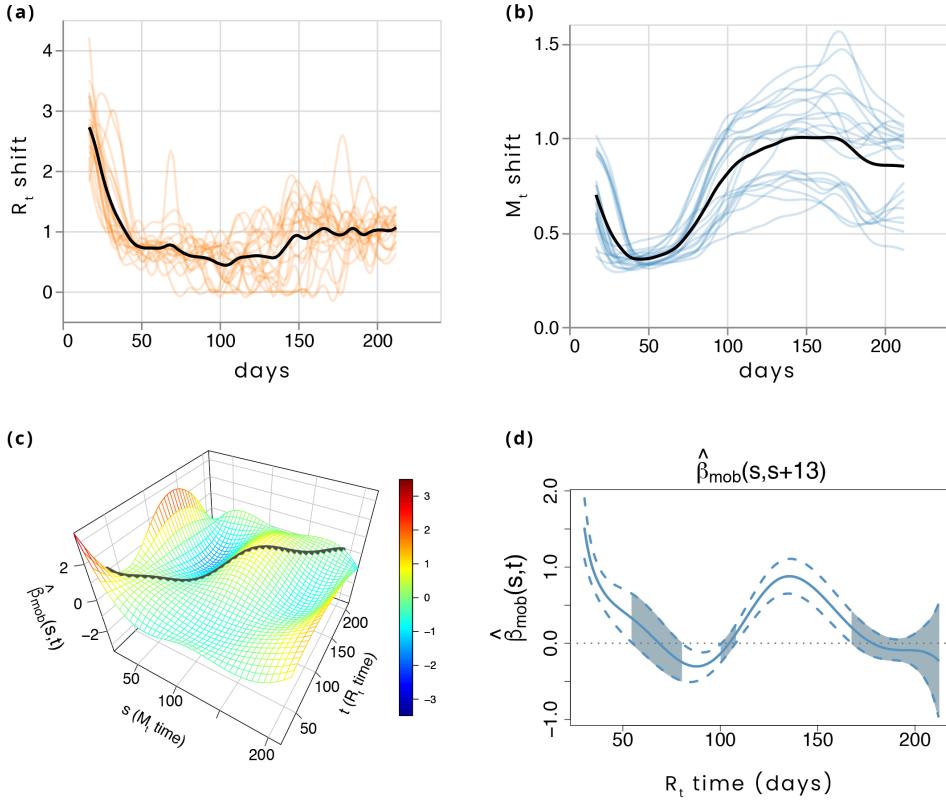
$$y(t) = \alpha(t) + \int \beta_{mob}(s, t) M_t(s) ds + \epsilon(t)$$

where  $y(t)$  is the shifted  $R_t$  curve,  $\alpha(t)$  is the intercept,  $\epsilon(t)$  is the model error, and  $M_t(s)$  is the shifted  $M_t$ . Mobility flows are integrated over time and the effect coefficient  $\beta_{mob}(s, t)$  is a surface that describes the association between  $M_t$  at time  $s$  and  $R_t$  at time  $t$ . Given the previously estimated 13 days lag between the two sets of curves, the portion of surface we are most interested in is the curve identified by  $\beta_{mob}(s, s+13)$ , which represents the effect of mobility at time  $s$  on transmission at time  $s+13$ . Panels (c) and (d) in Figure 3 show the estimated surface  $\hat{\beta}_{mob}(s, t)$  and the curve  $\hat{\beta}_{mob}(s, s+13)$  with a 95% confidence band<sup>69</sup>, respectively. This suggests a strong positive effect of mobility on transmissibility in March and April, and a second period of significant positive association in June (in these intervals of the domain the confidence bands are clearly above 0). The effect seems to be non-significant (confidence bands comprising 0) at the end of the time domain – which, because of the alignment process and the 13 days lag is ante-dated to the beginning of September.

The share of the variability in transmission explained by mobility on our data is high, approximately  $R^2 = 0.73$ .

To further validate our results we performed an additional analysis adding a scalar predictor ( $pc1$ ) to the regression model<sup>70</sup>. This is a composite control, obtained as the first principal component of a set of covariates that may affect the epidemic in addition to mobility – see Supplementary Figure S3 for a complete list. Introducing this control variable does not change our inference on the effect of mobility flows (see Supplementary Figure S3). The total  $R^2$  of the joint model is again 0.73, while the partial  $R^2$  of mobility and  $pc1$  are 0.24 and 0.01, respectively – this indicates how the effect of mobility is considerably stronger than the one of the control variable. As a second form of validation, we repeated our analysis using the 120 Italian provinces.

This finer spatial resolution increases our sample size, and at the same time produces more idiosyncratic curves.  $R_t$  is highly variable across provinces along the entire time domain considered.  $M_t$  is highly variable across provinces too, but especially during the summer, reflecting differences between tourist destinations and other areas, with increased and depressed mobility flows, respectively. Notwithstanding these differences, function-on-function regression produces results consistent with those at the level of regions.  $\hat{\beta}_{mob}(s, s + 13)$  displays a strong positive effect in March and April and a much weaker effect during the summer, with  $R^2 = 0.53$ .



**Figure 3:** Functional Data Analysis characterization of transmission and mobility. Panels (a) and (b) display  $R_t$  and  $M_t$  smoothed registered curves restricted to the intersection of their new time domains. Thick black lines are the functional averages of the sets of curves. After the alignment the time domain is expressed in days, since the horizontal translation changes the calendar dates for each region. Day 1 corresponds to March 9th. Panels in the bottom row show the estimated effect surface  $\hat{\beta}(s, t)$  and the curve  $\hat{\beta}(s, s + 13)$  obtained along it (black “cut” along the surface) with a 95% confidence band. The grey shaded areas represents the parts of the domain where the confidence bands contain 0.  $\hat{\beta}(s, s + 13)$  suggests a strong positive effect of mobility on transmission in March and April, and a second period of significant positive association in June. The share of the variability in transmission explained by mobility on our data is  $R^2 = 0.73$ .

## 5 Conclusions

Our combined analysis of mobility and epidemic data highlighted a striking relation between decreases in mobility flows and net reproduction number, in all Italian regions, in a time interval of approximately one week (from March 11th till March 16th) when the country transitioned between pre- and post-lockdown mobility regimes. During this week, the two curves gracefully overlap – until mobility stabilizes at about 40% of its pre-lockdown level. The two curves exhibit the same pattern in most of the regions and provinces (see Supplementary Figure S1), with minimal temporal lags. We call this phenomenon, represented schematically in Figure 2e, the “epi-mob” pattern. Mobility (the blue curve) depicts a switch between the two regimes, with an exponential fall from pre- to post-lockdown levels. The net reproduction number (the orange curve) displays a peak, with an exponential growth and fall, and overlaps with the switch during the latter.

Effectiveness of the lockdown for the containment of the epidemics is further substantiated by the positive association we found between delays in mobility reduction and total number of infected individuals per 100K inhabitants up to mid-May 2020. We can also quantify the time needed to “switch off” mobility in the country and the time needed to bring the net reproduction number below 1 (approximately 1 week).

Notice that  $R_t$  continues to slowly decrease during lockdown, and also that at the beginning of Phase 2 (lockdown exit), when mobility begins to raise again, for most of the regions  $R_t$  does not jump into a new uncontrolled growth for several months.

While we do not perform any analysis that may elucidate the role of other factors, this may be due to other non-pharmaceutical interventions taking hold in the country, including increased compliance with the use personal protective equipment and social distancing (compared to the pre-lockdown phase).

We have also shown, through functional data analysis, that the typical delay between the  $M_t$  peak and the  $R_t$  peak is around 2 weeks, both for regions and provinces. This allowed us to create a regression model which shows a strong positive effect of mobility on transmission in March, April, and June, both at regional and provincial level.

In conclusion, we believe that this study demonstrates the value of digital mobility data, a detailed proxy of human behavior available every day in real time, for refining our understanding of the dynamics of the epidemics, evaluating the effectiveness of policy choices, and monitoring the unfolding of the epidemics in the coming months.

**Acknowledgements** This work has been partially funded by the EU H2020 project SoBigData++ grant n. 871042. We thank the former Italian ministry of Innovation Paola Pisano, together with Paolo De Rosa and the colleagues of the working group “Big Data and AI for policy” of the Italian data-driven task force for the COVID-19 epidemics<sup>b</sup>, for supporting our work.

**Competing interest** The authors declare no competing interests.

**Code availability** The code to reproduce the analysis can be requested via email to luca.pappalardo@isti.cnr.it

---

<sup>b</sup><https://innovazione.gov.it/big-data-ai-for-policy>

**Author contributions** PC made data managing, data preparation and wrote the paper; LP coordinated the work, made plot design, and wrote the paper; SR and DF made data managing and visualization; TB made the functional data analysis experiments and plots; FG coordinated the work and wrote the paper; FC coordinated the functional data analysis experiments and wrote the paper; DP coordinated the work and wrote the paper; GG and SM provided feedback on the results and helped interpret them; Vodafone provided the data and made data preparation; Authors affiliated to ISS and FBK computed the  $R_t$  indicator.

## References

1. Perra, N. Non-pharmaceutical interventions during the covid-19 pandemic: a rapid review (2020). 2012.15230.
2. Bo, Y. *et al.* Effectiveness of non-pharmaceutical interventions on covid-19 transmission in 190 countries from 23 january to 13 april 2020. *International Journal of Infectious Diseases* **102**, 247 – 253 (2021). URL <http://www.sciencedirect.com/science/article/pii/S1201971220322700>.
3. Haug, N. *et al.* Ranking the effectiveness of worldwide covid-19 government interventions. *Nature human behaviour* 1–10 (2020).
4. Blondel, V. D., Decuyper, A. & Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Science* **4**, 10 (2015). URL <https://doi.org/10.1140/epjds/s13688-015-0046-0>.
5. Barbosa, H. *et al.* Human mobility: Models and applications. *Physics Reports* **734**, 1–74 (2018).
6. Luca, M., Barlacchi, G., Lepri, B. & Pappalardo, L. Deep learning for human mobility: a survey on data and models (2020). 2012.02825.
7. Giannotti, F. & Pedreschi, D. *Mobility, data mining and privacy: Geographic knowledge discovery* (Springer Science & Business Media, 2008).

8. Wang, Z., He, S. Y. & Leung, Y. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society* **11**, 141 – 155 (2018). URL <http://www.sciencedirect.com/science/article/pii/S2214367X17300224>.
9. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *nature* **453**, 779–782 (2008).
10. Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nature Communications* **6**, 8166 (2015). URL <https://doi.org/10.1038/ncomms9166>.
11. Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S. & Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour* **2**, 485–491 (2018).
12. Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
13. Pappalardo, L. *et al.* An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics* **2**, 75–92 (2016).
14. Eagle, N. & Pentland, A. S. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology* **63**, 1057–1066 (2009).
15. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
16. Barbosa, H. *et al.* Uncovering the socioeconomic facets of human mobility. *arXiv preprint arXiv:2012.00838* (2020).

17. Gabrielli, L., Furletti, B., Trasarti, R., Giannotti, F. & Pedreschi, D. City users' classification with mobile phone data. In *2015 IEEE International Conference on Big Data (Big Data)*, 1007–1012 (2015).
18. Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* **111**, 15888–15893 (2014). URL <https://www.pnas.org/content/111/45/15888>. <https://www.pnas.org/content/111/45/15888.full.pdf>.
19. Douglass, R. W., Meyer, D. A., Ram, M., Rideout, D. & Song, D. High resolution population estimates from telecommunications data. *EPJ Data Science* **4**, 4 (2015).
20. Hankaew, S. *et al.* Inferring and modeling migration flows using mobile phone network data. *IEEE Access* **7**, 164746–164758 (2019).
21. Balzotti, C., Bragagnini, A., Briani, M. & Cristiani, E. Understanding human mobility flows from aggregated mobile phone data. *IFAC-PapersOnLine* **51**, 25 – 30 (2018). URL <http://www.sciencedirect.com/science/article/pii/S2405896318307213>. 15th IFAC Symposium on Control in Transportation Systems CTS 2018.
22. Bonnel, P., Fekih, M. & Smoreda, Z. Origin-destination estimation using mobile network probe data. *Transportation Research Procedia* **32**, 69–81 (2018).

23. Liang, L., Shrestha, R., Ghosh, S. & Webb, P. Using mobile phone data helps estimate community-level food insecurity: Findings from a multi-year panel study in nepal. *PLOS ONE* **15**, 1–16 (2020). URL <https://doi.org/10.1371/journal.pone.0241791>.
24. Voukelatou, V. *et al.* Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics* 1–31 (2020).
25. Eagle, N., Macy, M. & Claxton, R. Network diversity and economic development. *Science* **328**, 1029–1031 (2010).
26. Frias-Martinez, V. & Virseda, J. On the relationship between socio-economic factors and cell phone usage. In *Proceedings of the fifth international conference on information and communication technologies and development*, 76–84 (2012).
27. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073–1076 (2015).
28. Lai, S. *et al.* Exploring the use of mobile phone data for national migration statistics. *Palgrave communications* **5**, 1–10 (2019).
29. Chi, G., Lin, F., Chi, G. & Blumenstock, J. A general approach to detecting migration events in digital trace data. *PloS one* **15**, e0239408 (2020).
30. Blumenstock, J. E. Inferring patterns of internal migration from mobile phone call records: evidence from rwanda. *Information Technology for Development* **18**, 107–125 (2012).

31. Hughes, C. *et al.* Inferring migrations: Traditional methods and new approaches based on mobile phone, social media, and other big data: Feasibility study on inferring (labour) mobility and migration in the european union from big data and social media data (2016).
32. de Montjoye, Y.-A. *et al.* On the privacy-conscious use of mobile phone data. *Scientific Data* **5**, 180286 (2018). URL <https://doi.org/10.1038/sdata.2018.286>.
33. Pellungrini, R., Pappalardo, L., Pratesi, F. & Monreale, A. A data mining approach to assess privacy risk in human mobility data. *ACM Trans. Intell. Syst. Technol.* **9** (2017). URL <https://doi.org/10.1145/3106774>.
34. Fiore, M. *et al.* Privacy in trajectory micro-data publishing : a survey. *arXiv: Cryptography and Security* (2019).
35. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports* **3** (2013).
36. Oliver, N. *et al.* Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Science Advances* **6** (2020). URL <https://advances.sciencemag.org/content/6/23/eabc0764>. <https://advances.sciencemag.org/content/6/23/eabc0764.full.pdf>.
37. Buckee, C. O. *et al.* Aggregated mobility data could help fight covid-19. *Science (New York, NY)* **368**, 145 (2020).
38. Finger, F. *et al.* Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proceedings of the National Academy of Sciences* **113**, 6421–6426 (2016).

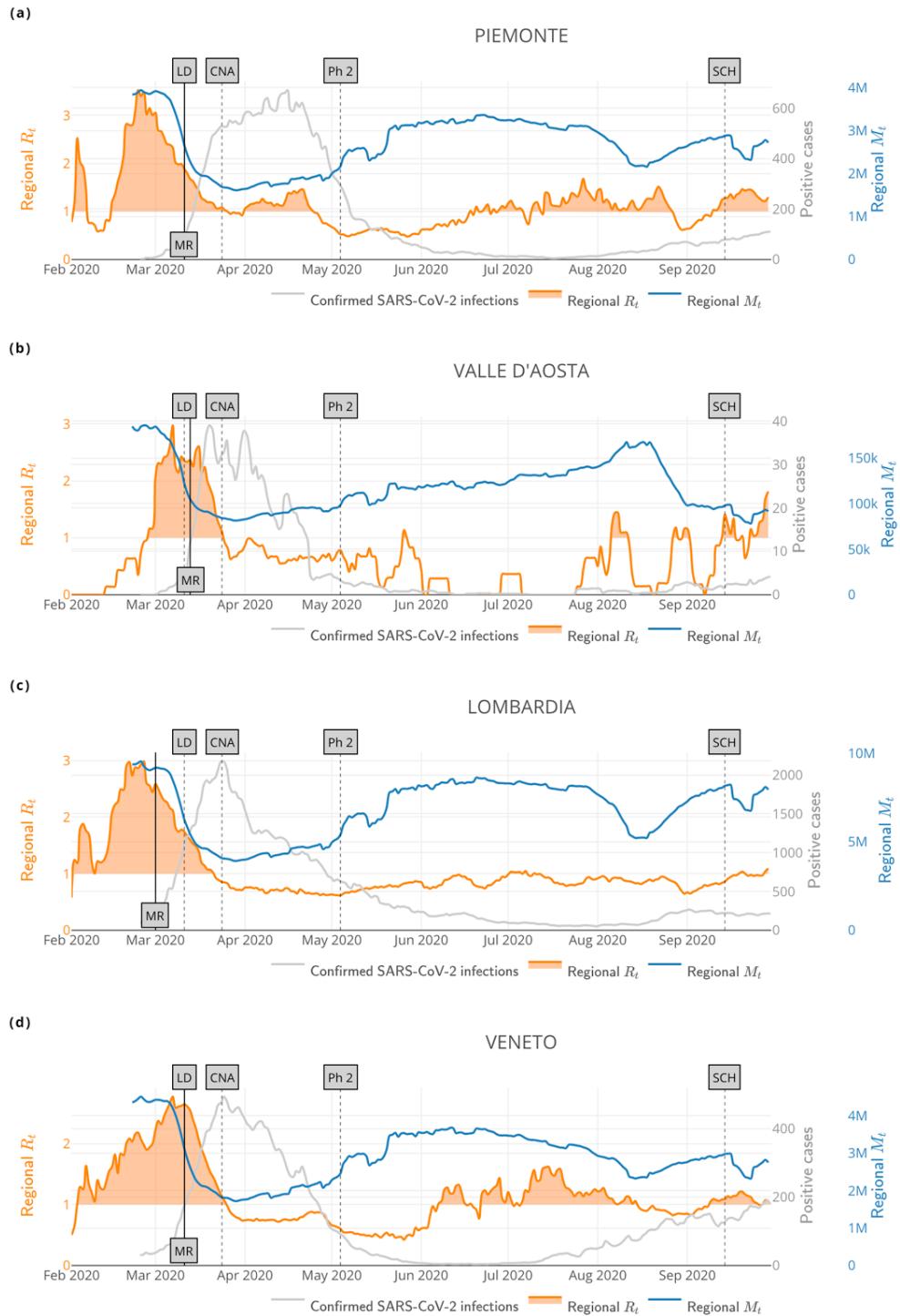
39. Tizzoni, M. *et al.* On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol* **10**, e1003716 (2014).
40. Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270 (2012).
41. Bengtsson, L. *et al.* Using mobile phone data to predict the spatial spread of cholera. *Scientific reports* **5**, 8923 (2015).
42. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**, 493–497 (2020).
43. Bonato, P. *et al.* Mobile phone data analytics against the covid-19 epidemics in italy: flow diversity and local job markets during the national lockdown (2020). 2004.11278.
44. Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S. & Colizza, V. Population mobility reductions during covid-19 epidemic in france under lockdown. *medRxiv* (2020).
45. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain covid-19 in china (2020).
46. Liautaud, P., Huybers, P. & Santillana, M. Fever and mobility data indicate social distancing has reduced incidence of communicable disease in the united states (2020). 2004.09911.
47. Badr, H. S. *et al.* Association between mobility patterns and covid-19 transmission in the usa: a mathematical modelling study. *The Lancet Infectious Diseases* (2020).
48. Coven, J. & Gupta, A. Disparities in mobility responses to covid-19. Tech. Rep., NYU Stern Working Paper (2020).

49. Gozzi, N. *et al.* Estimating the effect of social inequalities in the mitigation of covid-19 across communities in santiago de chile. *medRxiv* (2020).
50. Bakker, M., Berke, A., Groh, M., Pentland, A. & Moro, E. Effect of social distancing measures in the new york city metropolitan area. Tech. Rep., Massachusetts Institute of Technology (2020).
51. Jia, J. S. *et al.* Population flow drives spatio-temporal distribution of covid-19 in china. *Nature* 1–5 (2020).
52. Kissler, S. *et al.* Reductions in commuting mobility predict geographic differences in sars-cov-2 prevalence in new york city (2020).
53. Gibbs, H. *et al.* Changing travel patterns in china during the early stages of the covid-19 pandemic. *medRxiv* (2020).
54. Kang, Y. *et al.* Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic. *Scientific data* **7**, 1–13 (2020).
55. Pepe, E. *et al.* Covid-19 outbreak response, a dataset to assess mobility changes in italy following national lockdown. *Scientific data* **7**, 1–7 (2020).
56. Iqbal, M. S., Choudhury, C. F., Wang, P. & González, M. C. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies* **40**, 63–74 (2014).

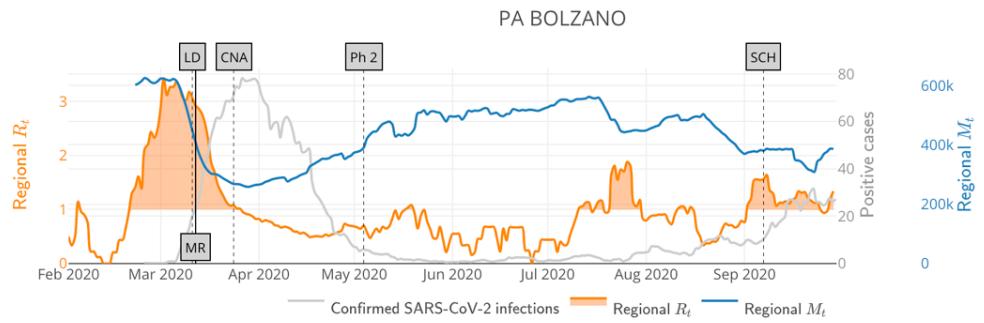
57. Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M. & Puchinger, J. Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies* **101**, 254–275 (2019).
58. Riccardo, F. *et al.* Epidemiological characteristics of covid-19 cases in italy and estimates of the reproductive numbers one month into the epidemic. *medRxiv* (2020).
59. Team, W. E. R. Ebola virus disease in west africa—the first 9 months of the epidemic and forward projections. *New England Journal of Medicine* **371**, 1481–1495 (2014).
60. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* **178**, 1505–1512 (2013).
61. Liu, Q.-H. *et al.* Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences* **115**, 12680–12685 (2018).
62. Nishiura, H., Linton, N. M. & Akhmetzhanov, A. R. Serial interval of novel coronavirus (covid-19) infections. *International journal of infectious diseases* (2020).
63. Wu, J. T. *et al.* Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, china. *Nature Medicine* **26**, 506–510 (2020).
64. Li, Q. *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).
65. Ramsay, J. O. & Silverman, B. W. *Functional data analysis* (Springer, 2005), 2 edn.

66. Craven, P. & Wahba, G. Smoothing noisy data with spline functions. *Numerische mathematik* **31**, 377–403 (1978).
67. Boschi, T., Chiaromonte, F., Secchi, P. & Li, B. Covariance based low-dimensional registration for function-on-function regression (2018).
68. Kokoszka, P. & Reimherr, M. *Introduction to functional data analysis* (CRC Press, 2017).
69. Goldsmith, J., Greven, S. & Crainiceanu, C. Corrected confidence bands for functional data using principal components. *Biometrics* **69**, 41–51 (2013).
70. Boschi, T., Di Iorio, J., Testa, L., Cremona, M. A. & Chiaromonte, F. The shapes of an epidemic: using functional data analysis to characterize covid-19 in italy. *arXiv preprint arXiv:2008.04700* (2020).

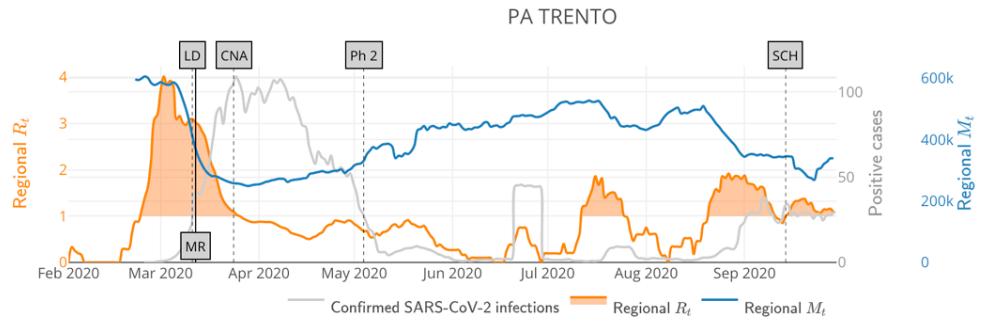
## **Supplementary Figures**



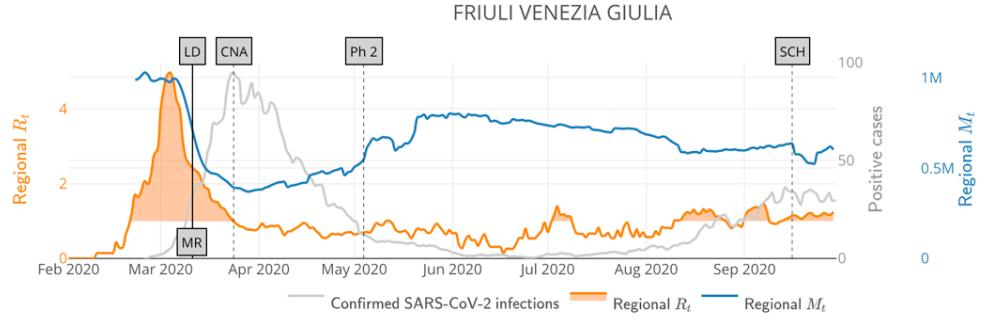
(e)



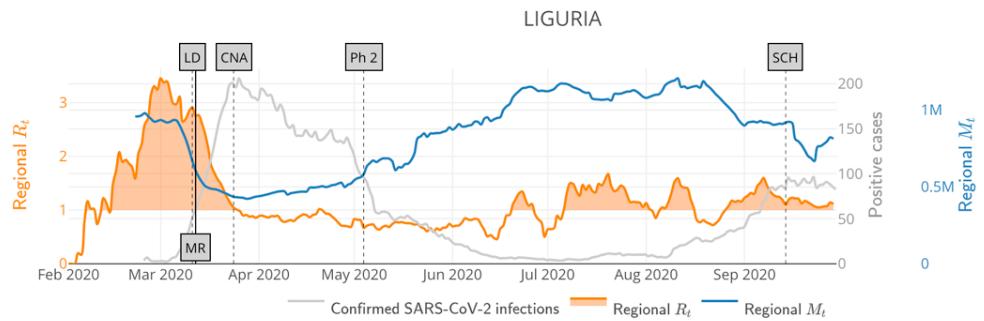
(f)

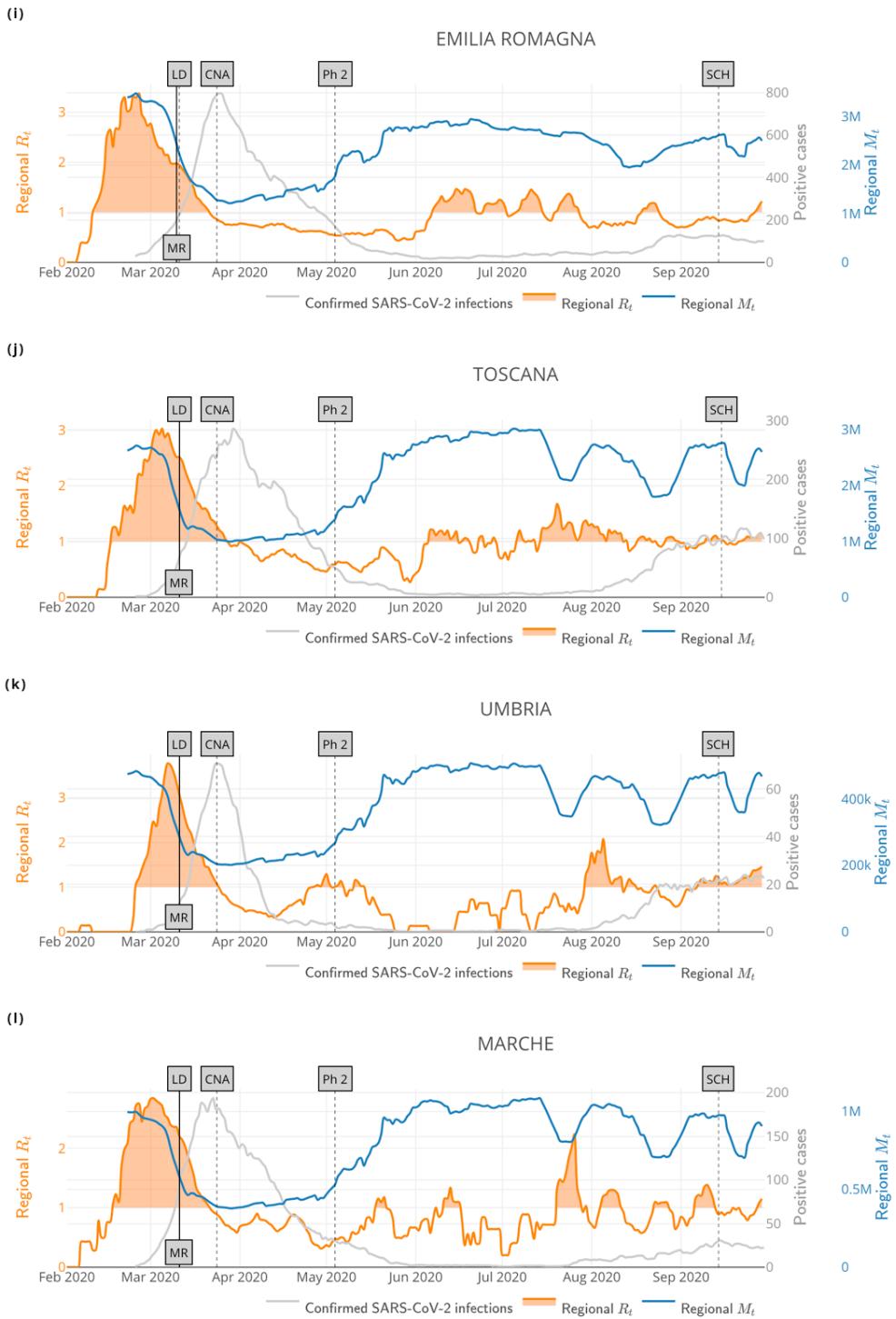


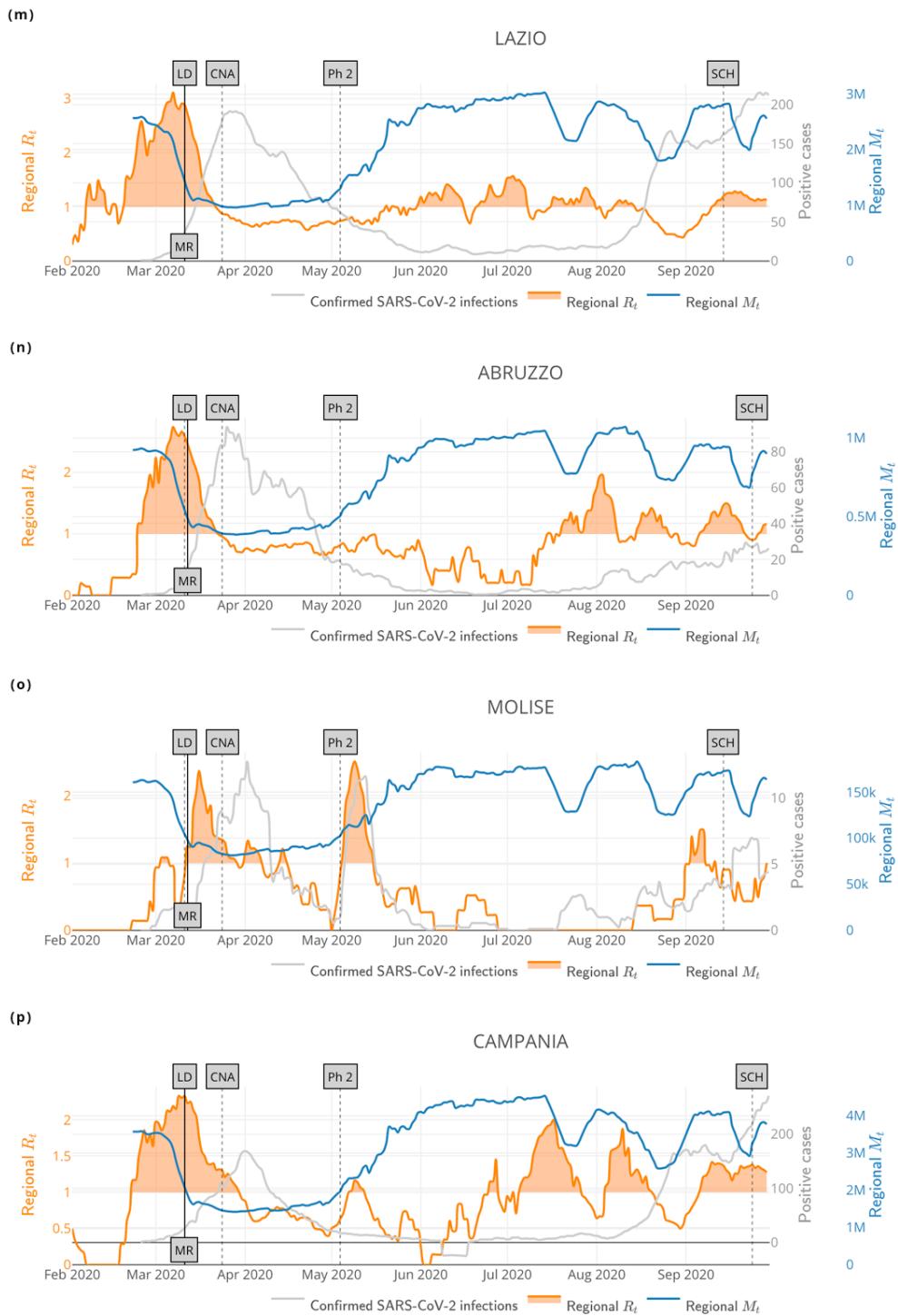
(g)

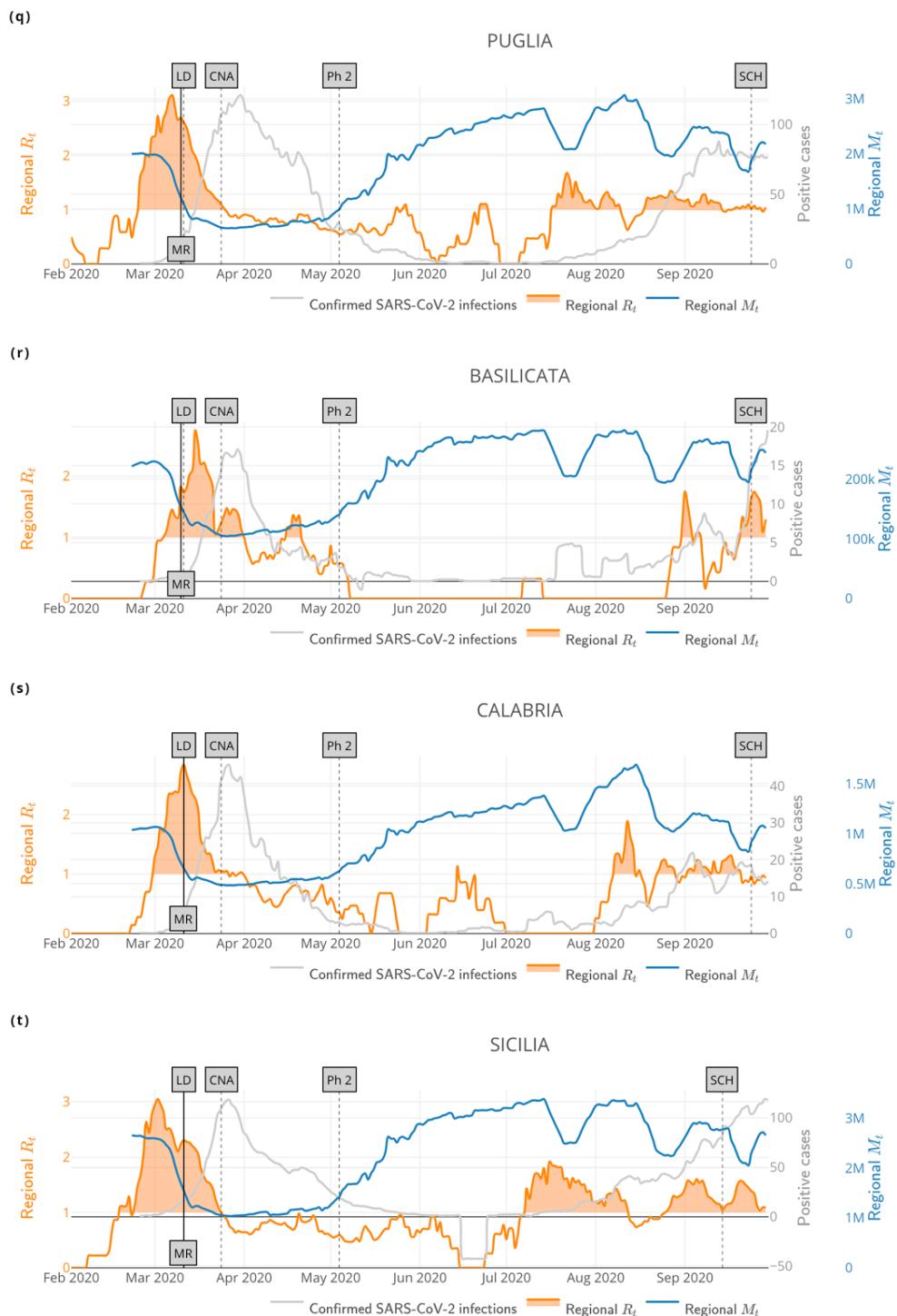


(h)









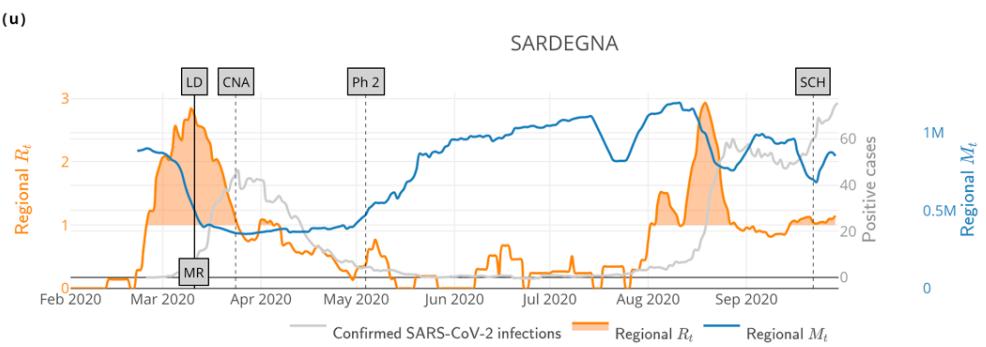


Figure 4: Line chart representations of the  $M_t$  (blue),  $R_t$  (orange) and the number of positive cases (grey) series for all the Italian regions. In each chart the series are normalized and projected on a dedicated axis for each curve, for an effective comparison of their evolution.

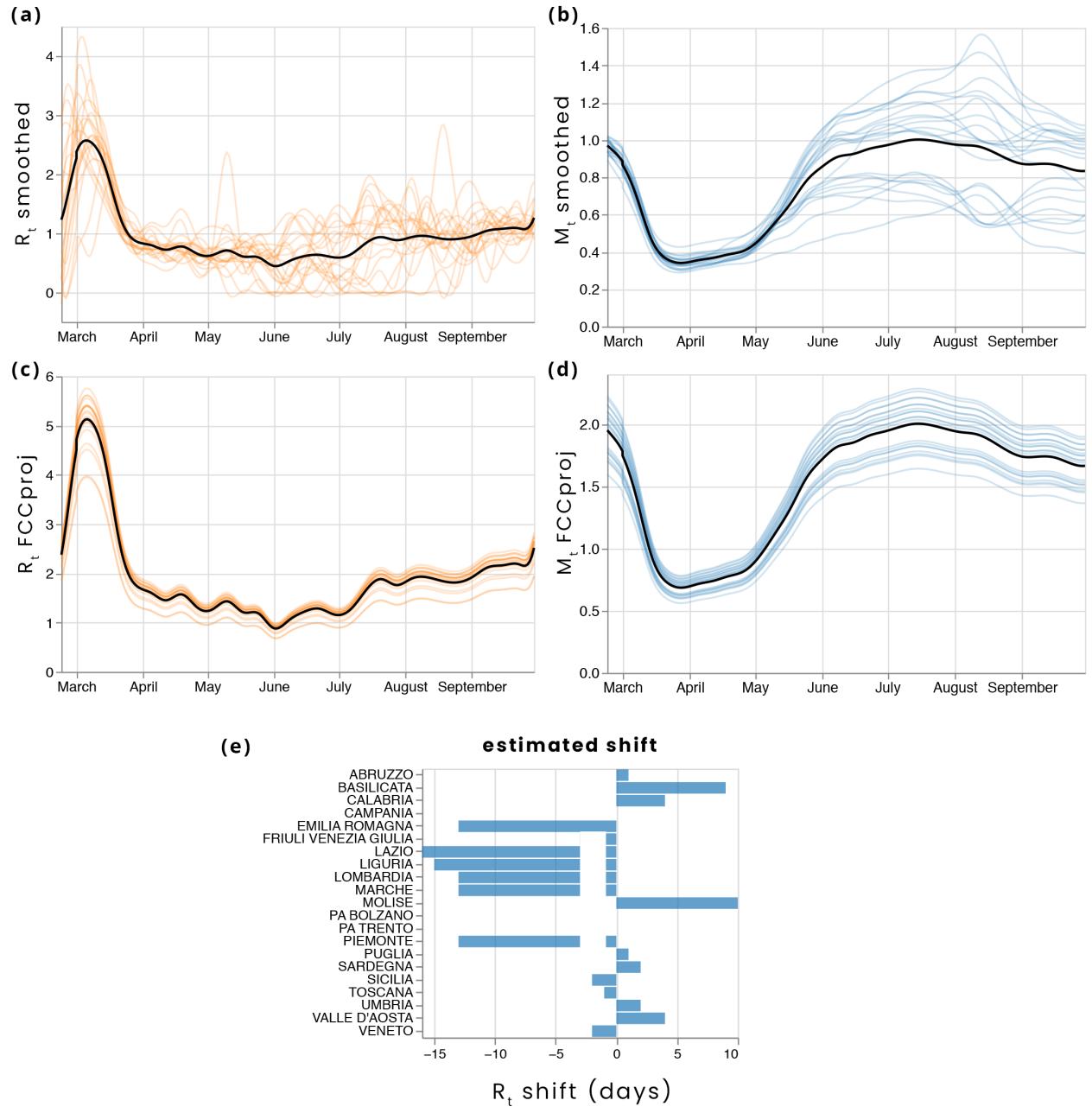


Figure 5: Panels (a) and (b) show smoothed  $R_t$  and  $M_t$  curves. Panels (c) and (d) display their leading covariance component projections. Thick black lines are the functional averages of the sets of curves. Panel (e) depicts the estimated horizontal shifts by the registration procedure.

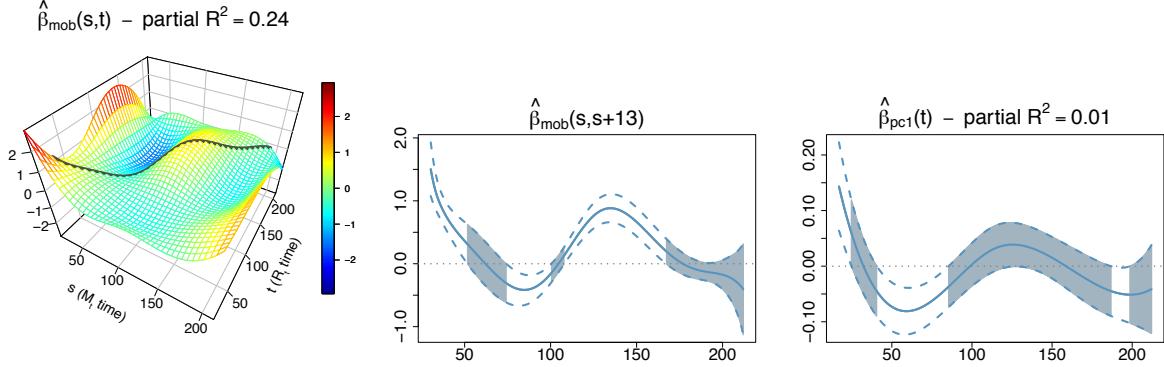


Figure 6: Results of the joint functional regression model. Transmission curves are regressed against mobility curves and the scalar variable  $pc1$ . This is a composite control, obtained as the first principal component of a set of covariates that may affect the epidemic in addition to mobility – specifically: adults per family doctor, average beds per hospital, average students per classroom, average employees per firm, and average members per household. The left and the center panels shows the estimated effect surface  $\hat{\beta}_{mob}(s, t)$  for mobility and the curve  $\hat{\beta}_{mob}(s, s + 13)$  obtained along it (black “cut” along the surface) with a 95% confidence band. The right panel shows the estimated  $pc1$  effect curve with a 95% confidence band. The gray shaded areas correspond to the parts of the domain where the confidence bands include the 0. The time along the axis is expressed in days. The starting date after the alignment process is postponed to March 9th.  $\hat{\beta}_{mob}(s, s + 13)$  suggests a strong positive effect of mobility on transmission in March and April, and a second period of significant positive association in June.  $\hat{\beta}_{pc1}(t)$  has an effect very close to 0 along with the all domain. The total  $R^2$  of the model is 0.73, while the partial  $R^2$  of mobility and  $pc1$  are 0.26 and 0.01, respectively.