Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Original Research Article

Mobility and the spatial spread of SARS-CoV-2 in Belgium

Michiel Rollier a,∗,1, Gisele H.B. Miranda a,c,1, Jenna Vergeynst a,b,1, Joris Meys a, Tijs W. Alleman a, the Belgian Collaborative Group on COVID-19 Hospital Surveillance d, Jan M. Baetens a

a KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
b BIOMATH, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
c DIvision of Computational Science and Technology, KTH Royal Institute of Technology, Tomtebodavägen 23A, Solna, 17165, Sweden
d Department of Epidemiology and Public Health, Sciensano, 1050 Brussels, Belgium

A R T I C L E   I N F O

Keywords:
covid-19
Epidemiology
Mobility
Time series analysis
Generalised additive mixed model

A B S T R A C T

We analyse and mutually compare time series of COVID-19-related data and mobility data across Belgium’s 43 arrondissements (NUTS 3). In this way, we reach three conclusions. First, we could detect a decrease in mobility during high-incidence stages of the pandemic. This is expressed as a sizeable change in the average amount of time spent outside one’s home arrondissement, investigated over five distinct periods, and in more detail using an inter-arrondissement “connectivity index” (CI). Second, we analyse spatio-temporal COVID-19-related hospitalisation time series, after smoothing them using a generalise additive mixed model (GAMM). We confirm that some arrondissements are ahead of others and morphologically dissimilar to others, in terms of epidemiological progression. The tools used to quantify this are time-lagged cross-correlation (TLCC) and dynamic time warping (DTW), respectively. Third, we demonstrate that an arrondissement’s CI with one of the three identified first-outbreak arrondissements is correlated to a substantial local excess mortality some five to six weeks after the first outbreak. More generally, we couple results leading to the first and second conclusion, in order to demonstrate an overall correlation between CI values on the one hand, and TLCC and DTW values on the other. We conclude that there is a strong correlation between physical movement of people and viral spread in the early stage of the SARS-CoV-2 epidemic in Belgium, though its strength weakens as the virus spreads.

1. Introduction

COVID-19 is a respiratory disease caused and spread by SARS-CoV-2, an infectious coronavirus. The first confirmed case in Belgium was a repatriated person from Wuhan [1], who tested positive on February 4, 2020 but did not spread the disease further (see Fig. 1 for a chronological overview). Returning vacationers from affected regions in Italy led to additional importations during the half-term holidays [2], and by early March 2020, disease transmission in Belgium was confirmed. On March 13, 2020, the Belgian government imposed the first measures to control the virus spread. On March 18, 2020, the measures were tightened to a lockdown. Two days later the Belgian borders were closed for non-essential travel, meaning that from that moment on cross-border mobility was severely restrained. Restrictions were gradually eased in May, June and July 2020. Upon the emergence of a second COVID-19 wave in September–October, measures were again restricted on October 19, 2020.

To inform policymakers about the forecasted evolution of the pandemic and the projected effects of containment measures, we developed a compartmental COVID-19 metapopulation model operational at the national scale [3] that tracks the number of individuals in 10 different epidemiological and clinical stages and 9 different age groups. It was developed in parallel with other researchers with slightly varying approaches [4,5]. The collective results of these efforts were finally bundled to obtain ensemble forecasts [6].

The aforementioned models work at the level of the entire Belgian population and hence do not capture the spatially heterogeneous nature of disease spread. However, spatially explicit phenomena such as human mobility have been shown to play an important role in the emergence of spatial disease dynamics in general [7–10]. For what concerns COVID-19 in particular, Refs. [11,12] have shown the importance of population flow on the dynamics of the emerging pandemic in China. On a smaller geographical scale and in a western context, Iacus et al. [13] demonstrated that the initial COVID-19 spread in France and

https://doi.org/10.1016/j.mbs.2022.108957
Received 30 March 2022; Received in revised form 10 November 2022; Accepted 19 December 2022
Available online 17 February 2023
0025-5564/© 2023 Elsevier Inc. All rights reserved.
Italy can be explained to a large extent by mobility between departments. A model that aims to correctly describes the spread of SARS-CoV-2 in large and diverse nations, therefore benefits from including a notion of population dynamics. Refs. [14–17] have developed such models for Spain, Brazil, France and the United Kingdom.

Adding a spatial component to our model [3] as well is both feasible and informative, as the relevant data are available and there is a demand for local insights and projections. Belgium, however, one would not typically classify as a large and diverse nation; with a road density of more than 500 km per 100 km² [19] and a 2020 population density of 374 inhabitants per km² [20], Belgium is one of the most connected and densely populated countries in the world. It is therefore not clear whether the relations between human mobility and SARS-CoV-2 spread observed in large countries, are also convincingly expressed for Belgium. In other words, it is not clear whether the inclusion of human mobility into a spatially explicit SARS-CoV-2 model is also necessary. The final objective of this study is to address that question.

The question is tackled by quantitatively addressing three sub-questions, all dealing with Belgian mobility and epidemiological time series in the year 2020. We first focus on mobility time series only and demonstrate that the average amount of time people spend outside their home arrondissement strongly declines during periods of elevated COVID-19 hospitalisation. This serves as a sanity check, and is primarily addressed to define the “connectivity index”. Second, we focus on COVID-19-related time series only, quantitatively discerning spatiotemporal structure in the spread of the disease. We do so by investigating, in terms of local COVID-19 hospitalisation time series, which arrondissements are “running ahead” of others, and which arrondissement pairs showed a morphologically similar COVID-19 evolution. We do so by applying time-lagged cross correlation and dynamic time warping, respectively. Third, these separate approaches are combined. In particular, we show that the strength of the connection to an early-outbreak arrondissement appears to predict significant excess mortality five to six weeks after the onset of the pandemic.

In general, our analysis indicates quantitatively that COVID-19-related time series of strongly connected arrondissements are on average more synchronised and morphologically similar than those of poorly connected arrondissements. This suggests that a comprehensive COVID-19 model for Belgium or similarly small countries may benefit from taking into account (fluctuating) mobility patterns. With such a model, the impact of regulating mobility during the early stages of a pandemic may be of particular interest.

Below, in Section 2, we first sketch the geographical situation in more detail and discuss the data used in our analysis. Section 3 present the methods for summarising and smoothing data, as well as both techniques for comparing time series. The closing Section 4 contains all results, i.e. the answers to the three sub-questions above, and provides a nuanced discussion and a concluding statement. This data analysis paper is followed by our modelling paper [21], which is currently under review.
of the population in any given arrondissement (private communication and Ref. [24]), we assume the data are representative of the entire population dynamics. The processed time series are expressed as the daily “staytime” \( P(t) \), denoting the total amount of time spent by all residents of arrondissement \( g \) in arrondissement \( h \) during the day corresponding with time \( t \). We will only use mobility quantities averaged over any of the four previously defined partial waves, and a prepandemic baseline.

3. Methods

3.1. Average daily mobility and connectivity index

To verify that mobility was severely affected during the pandemic (the first objective), we compare the average daily outward mobility during the ascending and descending phases of both the first and second 2020 COVID-19 waves with prepandemic mobility, per 100,000 inhabitants. We compute this as:

\[
(\text{average outward mobility})^g = \frac{1}{\Delta t} \frac{100000}{N^g} \sum_t \sum_{h \neq g} P^h(t),
\]

where superscript \( g \) denotes the considered (home) arrondissement, \( h \) the visited arrondissement, \( \Delta t \) the number of days in the considered period, and \( N^g \) the population of arrondissement \( g \). Note that we use superscripts for consistency with our other work [3,21].

We further define a mobility-based connectivity index (CI), inspired by Iacus et al. [13]:

\[
CI^g = \ln \left( \frac{1}{\Delta t_{\text{wave}}} \sum_{t \in \text{wave}} \sum_{r \in \text{wave}} (P^h(t) + P^h(t)) \right).
\]

This quantity expresses how well two arrondissements are connected over a certain period of time, irrespective of the direction of movement, i.e. \( CI^g = CI^h \) such that it is a property of the (unordered) pair \((g, h)\). The staytimes \( P^h(t) \) and \( P^h(t) \) in Eq. (2) are not normalised in order to arrive at an absolute measure of connectivity, rather than at an insight on how prone an individual in \( g \) is to visit \( h \). The natural logarithm is used for the sake of scaling, but other monotonically increasing functions may suffice as well, considering we will quantify our results with a rank correlation coefficient (see Section 3.4).

3.2. GAMM fitting and bootstrapping of COVID-19 time series

The unprocessed COVID-19 time series under consideration typically show high variability between days at the arrondissement level. This is partly explained by differences in reporting between week and weekend, and much of the additional variation comes from the inherently stochastic nature of (detecting) infections and hospitalisations. Rather than directly analysing noisy original data, we want to meaningfully compare trends between various arrondissements, and simultaneously provide an indication of uncertainty. An elegant way to achieve both goals is to model each time series using a Generalised Additive Mixed Model (GAMM) approach with a log link [25].

First, smoothing is established by modelling the expected count of events \( E(Y) \) per 100,000 inhabitants per day \( t \) as\
\[
E(Y_t) = \exp(a + s(X_t)\beta) + \epsilon_t,
\]

In this equation, \( a \) represents a general intercept. The vector \( s(X_t) \) is the \( r \)th row of a matrix containing the projection of time vector \( X \) on a second-order penalised B-spline basis. We chose the dimension of the spline basis to roughly match the number of weeks in the dataset, which allows for enough detail to incorporate the peaks in number of events, while the penalisation helps avoiding erratic fluctuations [26]. The dot product in the exponential is with \( \beta \), a vector of coefficients for the fixed effects.

If many people are hospitalised on day \( t \), fewer people remain to be hospitalised on day \( t+1 \), i.e. the data points in the time series are auto-correlated. Simultaneously, due to random events cancelling out, we are relatively more certain (in terms of signal-to-noise) of a high number of hospitalisations compared to a low number, i.e. depending on the statistical model, the data points may be overdispersed. The GAMM approach outperforms less advanced methods because it allows us to incorporate both autocorrelation and overdispersion. The residuals \( \epsilon_t \) are modelled according to a first-order autoregressive process (AR1) to account for auto-correlation between (subsequent) residuals. The estimation itself was performed using a quasi-likelihood method that assumes a linear relation between the mean and the variance to account for the documented overdispersion in COVID-19 transmission, as was done by an early Italian study by Scortichini et al. [27].

Fig. 3. illustrates the result of the GAMM approach for two Belgian arrondissements, Turnhout and Namur, on COVID-19 hospitalisation data. This plot also visualises the corresponding rolling average with a one-week window, illustrating the additional “smoothness” the GAMM approach provides, while keeping relevant epidemic trends.

Indicating uncertainty is established by exploiting the stochasticity inherent to the GAMM approach outlined above. For every original time series, we simulate not just 1 but 100 GAMM curves using a parametric bootstrap procedure based on the spline coefficients. In this procedure, we treat the coefficients as originating from a multivariate normal distribution, with the estimated value as mean and the estimated variance-covariance matrix of these coefficients as the variance structure. For every simulated time series a vector with coefficient values \( \beta' \) is randomly drawn from this distribution, and multiplied with the matrix \( s(X_t) \). Basically, original time series with a higher signal to noise ratio due to a limited number of daily counts will naturally produce more variability over the 100 GAMM fits, and hence more deviation over the resulting analysis values. We again refer to [25] for details, and to Fig. A.2 for an example of the resulting “spectrum” of fits.

3.3. Comparison between time series

Local COVID-19 time series generally resemble the epidemic behaviour at the national level, but we expect minor differences between arrondissements in amplitude and timing to be indicative of the geographical direction and intensity of the viral spread. More concretely, we anticipate that strongly connected arrondissements have near-synchronous time series with a highly analogous shape. We systematically compare hospitalisation time series for all unique arrondissement couples, for both the first and the second 2020 COVID-19 wave. We assess morphological similarities between smoothed time series using dynamic time warping (DTW). Difference in timing (lag) is computed using time-lagged cross-correlation (TLC). Both methods, which are briefly explained below, first impose standardisation to zero mean and unit standard deviation (“standard-score normalisation”). The significance
100 subtly varying GAMM fits are realised, allowing for a systematic quantification of the time series’ variability. The dashed curves represent the 7-day moving averages. The solid line is a single realisation following the Generalised Additive Mixed Model (GAMM) approach. Per arrondissement, 100 subtly varying GAMM fits are realised, allowing for a systematic quantification of the time series’ variability.

Time series for daily new COVID-19 hospitalisations per 100,000 inhabitants for the arrondissements Turnhout (dark red) and Namur (blue) for the entire first wave. The considered period for better comparison between distinct waves. The resulting time series for arrondissements g and h, which are here defined to vanish outside the considered time period. We are not interested in the value of \( C_{g,h}(k) \), which has little physical importance, but only in the horizontal position of the maximum, symbolised simply by

\[
\text{TLCC}^{gh} = \arg \max_k (C_{g,h}(k)).
\]

For instance, Fig. 3 suggests visually that for these two GAMM realisations, the peak of the daily new hospitalisations in Turnhout occurs earlier than in Namur. This is reflected by the maximum TLCC occurring at a lag of −5 days (Fig. 4, left), which allows us to conclude that Turnhout was about 5 days ahead of Namur for what concerns the time series for confirmed cases. Note however that the 99 other GAMM fits may generate different lag values, so the resulting value distribution’s standard deviation provides an indication for uncertainty. Additionally, note that this result, regardless of whether it is significantly non-zero, does not imply any type of causation.

DTW gauges morphological similarity by minimising the Euclidean distance between two standard-score normalised time series \( s^g(t) \) and \( s^h(t) \) of duration \( \Delta t \), through the deformation (remapping) of the time indices, after synchronisation based on TLCC lags [30]. The latter precondition is important in order to ascertain that the DTW value (often called DTW distance) does not intrinsically correlate with the TLCC value, hence safeguarding that both values communicate independent information. As the DTW distance scales with the number of analysed data points, we must normalise over the number of days \( \Delta t \) of the considered period for better comparison between distinct waves. Mathematically, we find a warping curve \( \phi(k) \) for which

\[
\phi(k) = \left( \phi_{g,k}(k), \phi_{h,k}(k) \right), \quad \text{with } \phi_{g,k}(k), \phi_{h,k}(k) \in \{1, \ldots, \Delta t\}
\]

which allows to minimise

\[
\text{DTW}^{gh} = d_\phi(s^g, s^h) = \frac{1}{\Delta t} \sum_{k=1}^{\Delta t} |\phi_{g,k}(k) - \phi_{h,k}(k)| m_g(k)/M_g,
\]

where \( m_g(k) \) is a per-step weighting coefficient and \( M_g \) is the corresponding normalisation constant. We refer to Giorgino (2009) [31] for details on the algorithm. Minimising the DTW value can be thought of as stretching or compressing the second series \( s^h \) with the aim of having it resemble as much as possible the first (reference) series \( s^g \). We are again not interested in the numerical value of this quantity per se, but rather in the comparison between distributions of such values for the 100 GAMM realisations of different arrondissement pairs, where lower DTW values indicate higher similarity. An illustration of the time axis deformation is shown in the right panel of Fig. 4.

### 3.4. Correlating mobility and COVID-19 time series

Third, we investigate the relation between mobility and SARS-COV-2 spread, by first looking at particular early-outbreak arrondissements and associated excess deaths, and then by generally investigating the correlation between CIs on the one hand, and DTW values and TLCC lags on the other.

We anticipate that Belgium follows the trends observed in Ref. [13]: a strong connection to an early-outbreak arrondissement on average foretells an increased excess mortality some weeks later. We aim to verify this by first identifying three early-outbreak arrondissements. For each of these arrondissements, we plot the associated first-wave ascending-phase CIs to all other arrondissements, and the average percentage of weekly excess deaths in these arrondissements, for every first-wave week since the start of the outbreak. The resulting time-dependent correlation between CIs and excess deaths (ED) for all \( G−1 = 42 \) arrondissements connected to arrondissement \( g \) is expressed as the Spearman’s rank correlation coefficient \( \rho \) [32]. This non-parametric
approach enables the detection of any monotonic relationship between both variables, which is desirable due to our agnosticism regarding the precise nature of the relation. Values closer to 1 (resp. −1) indicate stronger correlation (resp. anticorrelation).

Next, with all unique 903 CI values from Eq. (2) on the one hand, and all 903 TLCC lags and DTW values from Eqs. (4) and (5) on the other, we construct a number of scatter plots in which we quantify correlations between data pairs (CI$^{th}$, TLCC$^{th}$) and (CI$^{th}$, DTW$^{th}$). Note that we consider the absolute value for TLCC lags, because we are only interested in the magnitude of the time shift between the involved time series. We generally anticipate negative correlations: strongly connected arrondissements (high CI values) are expected to demonstrate small epidemiological time lags (small TLCC lags) and morphologically similar time series (small DTW values). We again calculate Spearman’s rank correlation coefficients. Values are calculated for the ascending, descending, and full period of both 2020 Belgian COVID-19 hospitalisation waves based on the respective GAMMs (100 per time series).

4. Results and discussion

4.1. Mobility changes during the 2020 COVID-19 waves

Confirming intuition, the average mobility outside the home arrondissement dropped substantially during the ascending and descending phases of both COVID-19 waves in comparison to the baseline mobility. This is shown clearly in the boxplots for the five chronologically ordered periods in Fig. 5, and in the geographically detailed thematic maps in Fig. 6. In particular, we observe that the change in outward mobility is less pronounced for the ascending parts compared to the ascending parts of both waves. This makes sense, considering that the lockdown period during both waves mainly coincides with their descending phases (see dashed double arrows in Fig. 1). Additionally we observe that overall average outward mobility was higher for the second wave compared to the first.

The connectivity indices defined in Eq. (2) demonstrate mobility changes in a more fine-grained fashion: (symmetric) heatmaps containing CIs are shown and compared for the two waves in Fig. 7. From this heatmap we infer that all but one arrondissements containing a province capital are relatively well connected to any other arrondissement according the CI metric, which is of course mainly due to the fact that such arrondissements typically have more inhabitants. Arlon is the only exception to this rule, which is expected because of its location, its small population of some 63,000, and the fact that a considerable number of its inhabitants normally commute abroad to the Grand Duchy of Luxembourg. This is in agreement with former studies on the general metropolitan connectivity in Belgium [33] and in line with the COVID-19-related comprehensive work by Islam et al. [34]. Comparing the first and second wave, the change in CI was relatively consistent between most arrondissement pairs: a small 5 to 10% increase during the second wave. Notable exceptions are connections with arrondissements in the province of West-Vlaanderen to the arrondissements Arlon, Bastogne, Virton, Veurne and Philippeville (over 15%). Despite highlighting the connectivity changes, the relevance of the actual CI values is limited due to the CI’s pragmatic definition. Still, they become meaningful in comparison to the corresponding DTW and TLCC values, to which we turn next.

4.2. Spatio-temporal dynamics of the 2020 COVID-19 waves

The heatmaps in Fig. 8 show the average time lag (in days) over all bootstrapped hospitalisation time series of both the first and second full COVID-19 wave, for each pair of arrondissements. The corresponding standard deviation can be found in Fig. A.3. From both figures, we infer for instance that the COVID-19 hospitalisation wave in Brussels-Capital during the first 2020 wave was about 1.6 ± 0.7 days (significantly) ahead of the one in the Antwerp arrondissement, and 0.5 ± 0.8 days (insignificantly) ahead of Diksmuide arrondissement.

According to Sciensano [18], the regions in the southwest of province Limburg and the so-called Borinage region are defined as the initial clusters in Belgium. These regions correspond with the arrondissements Tongeren and Hasselt (Limburg), and Mons (Borinage). Upon inspection of Fig. 8, this antecedence is most clearly visible for Mons, which has a negative lag on virtually all other arrondissements, and is a maximum of 6.3±1.3 days ahead compared to Namur. For Tongeren and Hasselt this is less convincingly so, arguably because the TLCC gauges the lag of the entire wave, while Sciensano identified initial clusters based on local index patients. We also notice that small arrondissements appear to have larger TLCC lags, which is presumably an artefact of the relatively large noise on these time series and the resulting GAMM fits. This is particularly visible for the arrondissements Diksmuide and Ieper during the first wave and translates to a very high TLCC lag standard deviation (see Figs. A.2 and A.3), and indicates that performing this analysis on even smaller geographical units would probably lead to meaningless results.

Interestingly, a clear distinction can be made between both COVID-19 waves when it comes to TLCC lags. Focusing on the larger geographical units, Brussels and Liège are now clearly ahead, and the province of West-Vlaanderen as a whole appears to lag behind. Generally, the lags observed during the second wave are overall larger than the ones observed during the first wave. This is seen in the boxplots in the bottom panel of Fig. 8, and quantified by a one-sided Wilcoxon signed-rank test [35], determining that the median of the differences is greater than zero with high confidence (p < 0.001). Additionally, the second-wave heatmap shows higher variability; this can be understood as the result of decreased mobility when compared to the ascending phase of the first wave (Fig. 5) — despite national homogenising during the summer and overall increased mobility during the entire wave. This suggests that a spatial analysis is especially feasible during times of low mobility, and already hints at a link between viral spread and population dynamics.

Similarly, Figs. 9 (resp. A.4) provide a comparison between arrondissements for both 2020 COVID-19 waves in terms of the average DTW distance (resp. standard deviations) over all GAMM realisations, between the corresponding hospitalisation time series.

For the first wave, the largest DTW distances are observed for low-population arrondissements. This can be understood as confirming our conjecture; these are of course also the least connected arrondissements (Fig. 7). However we must cautiously keep in mind that the GAMM procedure allows for high standard deviations when applied to noisy time series. For the second wave, generally, highly-populated and geographically close regions have time series that are similar in shape, as is seen by rectangular “clearings” around the diagonal in the heatmap. As was observed for TLCC lags, generally the DTW distances have grown for the second COVID-19 wave, and more variation between arrondissements is observed.
4.3. Connectivity and the initial spread of COVID-19 in Belgium

We consider Tongeren, Hasselt and Mons as early-outbreak arrondissements and plot the time-dependent correlation quantities in the top panel of Fig. 10. Clearly, the correlation coefficients for CI-versus-excess-death plots go up for arrondissements connected to Tongeren and to Hasselt (green and maroon curves) approximately two weeks after the defined start of the first COVID-19 wave. It remains quite constant, with a peak Spearman’s $\rho$ of resp. 0.63 (Tongeren) and 0.59 (Hasselt) some six weeks into the first wave. This indicates that arrondissements in our dataset that are well connected to these early-outbreak arrondissements, experience a higher excess mortality, which is according to expectation. This is best illustrated for Tongeren on April 15th in the map of Belgium in the bottom right panel of Fig. 10 — see the right panel of Fig. A.1 for the CI to Tongeren. The correlation coefficients time series for Mons remain noisy (blue curve). Despite being identified as an early-outbreak arrondissement by Sciensano [18], we do not see antecedent viral behaviour; nor in the TLCC lag analysis (Fig. 8), nor in this CI-versus-excess-mortality analysis. In any case, from May 2020 onward, the correlation coefficients become negligible and often even negative, demonstrating that connectivity index to the initial hubs is no longer correlated to local excess death: the virus spread has become nationally homogeneous.

The influence of connectivity to the French Haut-Rhin department on the initial spread disappeared 14 days after the first lockdown measures [13], which corresponds to the period between first symptoms and death [36]. The additional four weeks of delay in response of the virus spread to mobility changes in Belgium can be explained as follows: in contrast to the case of the Haut-Rhin department in France [13], no single big spreading event was documented in one of the initial clusters in Belgium, such that the effect on the percentage of excess deaths was more gradual. Furthermore, Belgium is a very connected country, and hence the difference between most and least connected arrondissements is smaller than in France.

4.4. Connectivity and COVID-19 dynamics

The general results showing correlations between CIs on the one hand, and DTW values or TLCC lags on the other, for first and second COVID-19 hospitalisation waves, are shown in Fig. 11. The highest Spearman’s rank correlation coefficients are retrieved for the DTW values, in particular for the first wave. The coefficients always point towards a negative correlation, albeit rather weak. The negative correlation coefficients endorse the conjecture that mobility is related to the spatial spread of COVID-19 strongly connected arrondissements will in general exhibit smaller delays (higher TLCC lags) and higher similarity in epidemiological progression (lower DTW values) when compared to poorly connected arrondissements. The scatter plots as well as the values of the correlation coefficients however suggest that the CI cannot be the only predictor for the velocity and morphology of COVID-19 spread. In order to complement the results shown in Fig. 5, the correlation analyses have been accomplished for the ascending and descending parts of the waves as well (Table A.2), resulting in lower correlation coefficients. This demonstrates that the analysis of the entire wave is needed to properly assess relations between morphology, synchronicity, and connectivity.

The results are in line with those from the study by Habib et al. [37], where a (non-linear) spatial linkage between COVID-19 and mobility is observed for Belgium as a whole. The correlations at a more fine-grained geographical level considered in this paper, are however weaker. This is in part due to noise of the involved data, but there are several other mobility-related factors that must have also played a role, and that could not be included in our analysis since relevant data are not available at the appropriate spatio-temporal resolution. Of course, the more fine-grained the geographical level, the more the latter are needed for a comprehensive analysis. These factors include the reason behind a mobility event (work, leisure, education, . . . ), the number of stops until the final destination of one mobility event, and so on. Furthermore, by working at NUTS 3 level, we ignored the role of short-distance mobility (mobility within an arrondissement) due to, for instance, local shopping, leisure and educational activities.
Fig. 8. Top: Average time-lagged cross-correlation lag computed between each pair of Belgian arrondissements from 100 GAMM fits on the hospitalisation time series, for first (a) and second (b) covid-19 waves. The colour scale indicates how many days (on average) the arrondissement indicated by the row header lags on the one indicated by the column header. Bottom: Identical TLCC information of all 903 unique arrondissement pairs, plotted in boxplots. There is a clear change between the first and the second wave, as shown by the bottom boxplot, with a median value of differences that is significantly greater than zero (Wilcoxon test \( p < 0.001 \)). Additional mean TLCC lag values as well as plots with standard deviations are found in Supplementary material.

Fig. 9. Top: Average normalised dynamic time warping distances computed between each unique pair of Belgian arrondissements from 100 GAMM fits on the hospitalisation time series, for first (a) and second (b) covid-19 waves. Bottom: identical DTW information, formatted as in Fig. 8. The median value of the differences between wave 1 and 2 is significantly greater than 0 (Wilcoxon test \( p < 0.001 \)).

Taking those short-distance movements into account, would probably allow us to identify a stronger relationship between mobility and covid-19 spread, as also Van De Vijver et al. (personal communication) pointed out that the spatial autocorrelation between covid-19 incidence dropped beyond 15 km for the first wave — while mostly exceeding 50 km during the second. This again indicates that the assumption of homogeneity within the arrondissements seems to be more valid at the start of the pandemic. Still, an even more fine-grained analysis would require the use of covid-19-related time series at municipal level, which are generally very noisy due to the relatively low number of hospitalisations per municipality, and hence complicating further statistical analysis without gaining much insight.

5. Conclusion

Results in this paper were presented in three stages that built up to the main conclusion: there is a strong correlation between physical movement of people and viral spread in the early stage of the SARS-CoV-2 epidemic in Belgium, which weakens once the virus has spread nationally.

We first confirmed that mobility between geographical regions at NUTS 3 level (arrondissements) in Belgium was reduced during the first and second wave of the covid-19 pandemic in 2020. Second, we quantified time lag and morphological similarity between local covid-19-related time series using dynamic time warping and time-lagged
cross-correlation. This approach proved to be meaningful and intuitive, provided we focus on full-wave hospitalisation time series, and we consider the large standard deviation over results associated with noisy data. Third, we assessed the strength of the relationship between the connectivity index of pairs of arrondissements on the one hand, and DTW or TLCC lags on the other. Particularly, we quantified the strength of the relationship between the connectivity to arrondissements that are affected first on the one hand, and local excess deaths on the other hand. We demonstrated a strong such correlation for the early-outbreak arrondissement Tongeren on the one hand, and a local excess mortality with a five to six week delay on the other hand. More generally, we observed a significantly nonzero but weak anticorrelation, notably for DTW distances. This confirms that, in our dataset, strongly connected arrondissements exhibit morphologically similar hospitalisation time series, that are (on average) roughly synchronised. However, other factors beyond the control of our analysis appear to cloud a clean correlation.

The techniques developed in and conclusions drawn from this research demonstrate that a spatio-temporal data analysis of mobility and epidemiological data at NUTS 3 level in Belgium is feasible and informative. This motivates data analysis for other sociological and demographic aspects of society that may be employed as a metric for the SARS-CoV-2 pandemic. Moreover, the conclusions imply that a model for COVID-19 in Belgium may benefit from including a notion of mobility, especially when modelling the early stages of the SARS-CoV-2 pandemic, before the virus has spread homogeneously throughout the country. We have therefore set up a spatially explicit model based on mobility data within Belgium [21].
Table A.1
43 Belgian arrondissements (NUTS 3) ordered by systematic identification (first two digits of NIS code) and blocked per province (NUTS 2), with population size, area and population density. Arrondissements containing a (province) capital are in boldface. The NUTS 1 regions are indicated by F (Flanders, Dutch names), W (Wallonia, French names) or B (Brussels-Capital Region, English name).

| NIS | Arrondissement (NUTS 3) | Pop. (NUTS 3) | Area (km$^2$) | Density (km$^{-2}$) | Province (NUTS 2) | Region (NUTS 1) |
|-----|-------------------------|----------------|---------------|---------------------|--------------------|----------------|
| 11  | Antwerpen               | 1 057 736      | 1 004         | 1 053               | Antwerpen          | F              |
| 12  | Mechelen                | 347 125        | 511           | 678                 | Antwerpen          | F              |
| 13  | Turnhout                | 464 869        | 1 360         | 341                 | Antwerpen          | F              |
| 21  | Brussels-Capital        | 1 218 255      | 362           | 3 500               | N/A                | F              |
| 23  | Halle-Vilvoorde         | 643 766        | 949           | 678                 | Vlaams-Brabant     | F              |
| 24  | Leuven                  | 512 077        | 1 169         | 437                 | Vlaams-Brabant     | F              |
| 25  | Nivelles                | 406 019        | 1 097         | 370                 | Brabant Wallon     | W              |
| 31  | Brugge                  | 282 745        | 673           | 419                 | West-Vlaanderen    | F              |
| 32  | Diest                   | 51 696         | 365           | 141                 | West-Vlaanderen    | F              |
| 33  | Turnhout                | 106 570        | 553           | 192                 | West-Vlaanderen    | F              |
| 34  | Korrigijk               | 292 493        | 406           | 720                 | West-Vlaanderen    | F              |
| 35  | Oostende                | 157 780        | 304           | 518                 | West-Vlaanderen    | F              |
| 36  | Roeselare               | 154 494        | 273           | 564                 | West-Vlaanderen    | F              |
| 37  | Tielrode                | 93 428         | 331           | 281                 | West-Vlaanderen    | F              |
| 38  | Verviers                | 61 739         | 288           | 214                 | West-Vlaanderen    | F              |
| 41  | Aalst                   | 293 650        | 472           | 620                 | Oost-Vlaanderen    | F              |
| 42  | Dendermonde             | 202 411        | 346           | 584                 | Oost-Vlaanderen    | F              |
| 43  | Eeklo                   | 85 692         | 335           | 255                 | Oost-Vlaanderen    | F              |
| 44  | Gent                    | 564 042        | 949           | 593                 | Oost-Vlaanderen    | F              |
| 45  | Oudenaarde              | 124 610        | 422           | 294                 | Oost-Vlaanderen    | F              |
| 46  | Sint-Niklaas            | 254 850        | 479           | 531                 | Oost-Vlaanderen    | F              |
| 51  | Ath                     | 128 468        | 671           | 191                 | Hainaut            | W              |
| 52  | Charleroi               | 396 962        | 475           | 834                 | Hainaut            | W              |
| 53  | Mons                    | 259 237        | 588           | 440                 | Hainaut            | W              |
| 55  | Soignies                | 105 179        | 357           | 294                 | Hainaut            | W              |
| 56  | Thuin                   | 91 725         | 785           | 116                 | Hainaut            | W              |
| 57  | Tournai-Mouscron        | 223 799        | 714           | 313                 | Hainaut            | W              |
| 58  | La Louvière             | 141 470        | 219           | 644                 | Hainaut            | W              |
| 61  | Huy                     | 113 869        | 661           | 172                 | Liège              | W              |
| 62  | Liège                   | 620 765        | 795           | 786                 | Liège              | W              |
| 63  | Verviers                | 288 277        | 2009          | 143                 | Liège              | W              |
| 64  | Waregem                 | 81 889         | 390           | 209                 | Liège              | W              |
| 71  | Hasselt                 | 420 312        | 883           | 475                 | Limburg            | F              |
| 72  | Maaseik                 | 252 115        | 910           | 276                 | Limburg            | F              |
| 73  | Tongeren                | 204 943        | 633           | 323                 | Limburg            | F              |
| 81  | Arlon                   | 62 996         | 318           | 197                 | Luxembourg         | W              |
| 82  | Bastogne                | 490 83        | 1046          | 46                  | Luxembourg         | W              |
| 83  | Marche-en-Famenne        | 56 771         | 958           | 59                  | Luxembourg         | W              |
| 84  | Neufchâteau             | 63 763         | 1358          | 46                  | Luxembourg         | W              |
| 85  | Virton                  | 54 139         | 777           | 69                  | Luxembourg         | W              |
| 91  | Dinant                  | 111 286        | 1596          | 69                  | Namur              | W              |
| 92  | Namur                   | 318 231        | 1167          | 272                 | Namur              | W              |
| 93  | Philippeville           | 66 315         | 910           | 72                  | Namur              | W              |

Fig. A.1. Left: Heatmap showing the second-wave CIs, defined in Eq. (2). Right: geographical detail of CIs to Tongeren during the ascending phase of the first wave, on the same colour scale, with an indication of arrondissements Hasselt and Liège.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors wish to express their gratitude to Proximus, Belgium’s leading telecom operator, for their generous commitment in the context of the Task Force ‘Data & Technology against Corona’; in particular for providing daily detailed mobility data.

Funding

This work was supported by the UGent Special Research Fund, Belgium, by the Research Foundation Flanders (FWO), Belgium, project numbers G0G2920 and G0G9820, and by VZW 100 km Dodentocht Kade, Belgium through the organisation of the 2020 100 km COVID-Challenge. Further, the computational resources and services used in this work were also provided by the VSC (Flemish Supercomputer Center), funded by FWO, Belgium and the Flemish Government, Belgium. The funding sources played no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

Supplementary information

This article is associated with a supplementary document containing additional information on the Belgian geography, a more in-depth discussion of the GAMM fitting, and with additional results. It is currently included as an appendix.

Appendix. Supplementary material

A.1. Additional geographical information

Table A.1 contains all relevant demographic and geographical information on the 43 Belgian arrondissements. The left-hand side of Fig. A.1 demonstrates the symmetric matrix CI on the 43 Belgian arrondissements. The right-hand side of Fig. A.1 shows the CI to Tongeren during the ascending phase of the first wave (compare this to Fig. 10). Clearly, geographical distance is related to the CI.

A.2. Details of the time series GAMM fitting

Mathematical construction. Our approach for data smoothing and uncertainty indication makes use of GAMM fitting and bootstrapping [25], described in the main text. We carried out all calculations with the mgcv package [25] in the statistical software R [38]. Note, for completeness, that due to the log transformation the model fit becomes unstable if there are long periods without events. To increase numerical stability, we simply add 1 to the data prior to fitting in these cases. This correction is subsequently subtracted from the resulting predictions.

Assumptions and weaknesses. We use GAMMs as a more advanced smoothing technique and as a means to indicate uncertainty. Contrary to moving averages and other local smoothing techniques, they allow us to incorporate both autocorrelation and overdispersion. The resulting smooth curves are in general less sensitive to erratic fluctuations in the data, while still sufficiently flexible to describe general trends (and differences between these) in a time series of number of events [25,26].

The GAMM framework also offers a more formal estimation of the uncertainty on the parameters. This allows for a computationally efficient method to construct bootstrap samples from these general trends, which are in turn used to formulate a spectrum of slightly deviating results. A possible weakness lies in the fact that this estimate of uncertainty relies heavily on a number of assumptions. First of all, we assume that the β coefficients follow a multivariate normal distribution. Further, the Laplace approximation performs well for a sufficiently large sample size [39]. We assume that the amount of data used is sufficient to expect little deviation from this assumption.

Second, we assume a linear relationship between the mean and variance of predictions. This approach, often referred to as quasi-Poisson, has been used in numerous other analyses (e.g. Refs. [40–42]). Yet, assuming a quadratic relationship between mean and variance would be more equivalent to the negative binomial distribution assumed by Endo et al. [43]. On the other hand, such an approach would give larger values less weight in the fit compared to the quasi-Poisson method [44]. Comparison of both methods for a selection of arrondissements showed that assuming a quadratic relationship would lead to a systematic underestimation of the peak height.

By analysing the squared deviation from weekly averages, we concluded that the linear assumption could be defended. Only for the few cases with a very large number of events, variance would be underestimated. In the majority of cases, the linear relation would slightly overestimate the variance, making the approach more conservative.

A.3. Additional DTW and TLCC results

The heatmaps and boxplots in Fig. A.3, associated with Fig. 8, communicate the standard deviations of TLCC lags, taken over 100 slightly different GAMM realisations of the hospitalisation time series of 903 unique arrondissement pairs, for the entire first covid-19 wave
Fig. A.3. Formatting as in Fig. 8. Top: Standard deviation of all 100 normalised dynamic time warping distances computed between each unique pair of Belgian arrondissements from 100 GAMM fits on the hospitalisation time series, for full first (a) and second (b) COVID-19 waves. Bottom: Boxplots for the same 903 unique arrondissement pairs, demonstrating visually the difference in distributions (Wilcoxon test $p < 0.001$).

Fig. A.4. Formatting as in Fig. 9, again communicating standard deviations as in Fig. A.3, but now for DTW values of arrondissement pairs. Wilcoxon test $p < 0.001$.

Table A.2

| Spearman’s rank correlation coefficient | Wave 1 | Wave 2 |
|----------------------------------------|--------|--------|
|                                        | TLCC   | DTW    | TLCC   | DTW    |
| Full                                   | -0.14(4) | -0.25(4) | -0.15(4) | -0.31(3) |
| Ascending                              | -0.14(4) | -0.15(5) | -0.11(5) | -0.16(4) |
| Descending                             | -0.14(4) | -0.19(3) | -0.11(6) | -0.31(4) |

in Belgium. Fig. A.4, associated with Fig. 9, shows the same, but for DTW values. A comprehensive table for all results is provided in Table A.2.

References

[1] F.P.S. Health, One repatriated Belgian has tested positive for the novel coronavirus, 2020, URL https://web.archive.org/web/20200406101412/https://www.info-coronavirus.be/en/news/one-repatriated-belgian-has-tested-positive-for-the-novel-coronavirus/, Accessed 2020-02-04.

[2] F.P.S. Health, Returns from Italy push COVID-19 tally higher, 2020, URL https://www.vrt.be/vrtnws/en/2020/03/02/returns-from-italy-push-covid-19-tally-higher/, Accessed 2020-03-02.

[3] T.W. Alleman, J. Vergeynst, L.D. Visscher, M. Rollier, E. Torfs, the Belgian Collaborative Group on COVID-19 Hospital Surveillance, I. Nopens, J.M. Baetens, Assessing the effects of non-pharmaceutical interventions on SARS-CoV-2 spread in Belgium by means of a compartmental, age-stratified, extended SEIQRD model and public mobility data, Epidemics 37 (2021) 100505.

[4] S. Abrams, J. Wambua, E. Santermans, L. Willems, E. Kuylen, P. Coletti, P. Libin, C. Faes, O. Petrof, S.A. Herzig, P. Beutels, N. Hens, Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model.
model and studying its implied future trajectories, Epidemics 35 (2021) 100449, http://dx.doi.org/10.1016/j.epidem.2021.100449.

5. N. Franco, COVID-19 Belgium: Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts, Epidemics 37 (2021) 100490, http://dx.doi.org/10.1016/j.epidem.2021.100490.

6. RESTORE, RESTORE report B: Long-term scenarios for the number of new hospitalizations during the belgian COVID-19 epidemic, 2021, URL https://covid-en.wissenschaft.github.io/assets/restore/report_v8_0.pdf, Accessed 2021-06-29.

7. S. Changrangenmug, D.J. Bicout, C. Modchang, How the individual human mobility spatio-temporally shapes the disease transmission dynamics, Sci. Rep. 10 (1) (2020) 1–13, http://dx.doi.org/10.1038/s41598-020-68230-9.

8. S. Merler, M. Ajelli, The role of population heterogeneity and human mobility in the spread of pandemic influenza, Proc. R. Soc. B: Biol. Sci. 277 (1681) (2010) 557–565, http://dx.doi.org/10.1098/rspb.2009.1605.

9. A. Wesolowski, N. Eagle, A.J. Tatem, D.L. Smith, A.M. Noor, R.W. Snow, C.O. Buckee, Quantifying the impact of human mobility on malaria, Science 338 (6104) (2012) 267–270, http://dx.doi.org/10.1126/science.1235467.

10. A. Wesolowski, T. Qureshi, M.F. Boni, P.R. Sundey, M.A. Johansson, S.B. Rasheed, K. Engo-Monsen, C.O. Buckee, Impact of human mobility on the emergence of dengue epidemics in Pakistan, Proc. Natl. Acad. Sci. USA 112 (38) (2015) 11887–11892, http://dx.doi.org/10.1073/pnas.1504941112.

11. Z. Du, L. Wang, S. Cauchemez, X. Xu, X. Wang, B.J. Cowling, L.A. Meyers, Transportation of coronavirus disease from Wuhan to other cities in China, Emerg. Infect. Diseases 26 (5) (2020) 1049–1052, http://dx.doi.org/10.3201/ eid2605.200146.

12. J. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, N. Christakis, Population flow drives spatio-temporal distribution of COVID-19 in China, Nature (2020) http://dx.doi.org/10.1038/s41586-020-2284-y.

13. S.M. Iacus, C. Santamaria, F. Sera, S. Spyratos, D. Tarchi, M. Vespe, How the spread of pandemic influenza, Proc. R. Soc. B: Biol. Sci. 277 (1681) (2010) 557–565, http://dx.doi.org/10.1098/rspb.2009.1605.

14. A. Arenas, W. Cota, S.C. Ferreira, Metapopulation modeling of COVID-19 advancing into the countryside: an analysis of mitigation strategies for Brazil, MedRxiv (2020) http://dx.doi.org/10.1101/2020.05.06.20093492, 2020.05.06.20093492.

15. N. Islam, V.M. Shkolnikov, R.I. Acosta, I. Klimkin, I. Kawachi, R.A. Izarzry, G. Alicantro, K. Khunti, T. Yates, D.A. Janov, M. White, S. Lewington, B. Lacey, Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries, BMJ 373 (2021) http://dx.doi.org/10.1136/BMJN137.

16. F. Wilcoxson, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (1945) 80–83, Accessed 2022-11-09.

17. S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application, Ann. Intern. Med. 172 (9) (2020) 577–582, http://dx.doi.org/10.7326/M20-0504.

18. Y. Habib, E. Xia, S.H. Hashmi, Z. Fareed, Non-linear spatial linkage between COVID-19 pandemic and mobility in ten countries: A lesson for future wave, J. Infect. Public Health 14 (10) (2021) 1411–1426, http://dx.doi.org/10.1016/J.JIHPH.2021.08.008.

19. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021, https://www.R-project.org/.

20. G. Kauermann, K. Boveng, Quasi-Poisson vs. Negative binomial regression: how should we model overdispersed count data? Ecology 88 (11) (2007) 2766–2772, http://dx.doi.org/10.1890/07-0043.1.

21. Y. Habib, E. Xia, S.H. Hashmi, Z. Fareed, Non-linear spatial linkage between COVID-19 pandemic and mobility in ten countries: A lesson for future wave, J. Infect. Public Health 14 (10) (2021) 1411–1426, http://dx.doi.org/10.1016/J.JIHPH.2021.08.008.

22. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021, https://www.R-project.org/.

23. G. Kauermann, K. Boveng, Quasi-Poisson vs. Negative binomial regression: how should we model overdispersed count data? Ecology 88 (11) (2007) 2766–2772, http://dx.doi.org/10.1890/07-0043.1.