



# Mobility functional areas and COVID-19 spread

Stefano Maria Iacus<sup>1</sup> · Carlos Santamaria<sup>1</sup> · Francesco Sermi<sup>1</sup> · Spyridon Spyratos<sup>1</sup> · Dario Tarchi<sup>1</sup> · Michele Vespe<sup>1</sup>

© The Author(s) 2021

## Abstract

This work introduces a new concept of functional areas called Mobility Functional Areas (MFAs), i.e., the geographic zones highly interconnected according to the analysis of mobile positioning data. The MFAs do not coincide necessarily with administrative borders as they are built observing natural human mobility and, therefore, they can be used to inform, in a bottom-up approach, local transportation, spatial planning, health and economic policies. After presenting the methodology behind the MFAs, this study focuses on the link between the COVID-19 pandemic and the MFAs in Austria. It emerges that the MFAs registered an average number of infections statistically larger than the areas in the rest of the country, suggesting the usefulness of the MFAs in the context of targeted re-escalation policy responses to this health crisis. The MFAs dataset is openly available to other scholars for further analyses.

**Keywords** COVID-19 · Coronavirus · Mobile positioning data · Mobility functional areas

## Introduction

In April 2020, the European Commission (EC) asked European Mobile Network Operators (MNOs) to share fully anonymised and aggregated mobility data in order to support the fight against COVID-19 (European 2020, 2020) with data driven evidence.

The value of mobile positioning *personal data* to describe human mobility has been already explored in literature (Csáji et al. 2013) and its potential in epidemiology demonstrated (Wesolowski et al. 2012; Jia et al. 2020; Wu et al. 2020; Kraemer et al. 2020). The new initiative between the Commission and the European MNOs relies on the effectiveness of using *fully anonymised and aggregated mobile positioning data* in compliance with ‘Guidelines on the use of location data and contact tracing tools in the context of the COVID-19 outbreak’ by the European Data Protection Board (EDPB 2020).

This work makes use of mobile data to introduce a new concept of ‘functional areas’ called *Mobility Functional Areas* or MFAs. The concept of functional areas has a long tradition in settlement geography, urban planning and policy making (Ball 1980; Van der

---

✉ Stefano Maria Iacus  
stefano.iacus@ec.europa.eu

<sup>1</sup> European Commission, Joint Research Centre, Via Enrico Fermi 2749, 21027 Ispra (VA), Italy

Laan 1998; Casado-Díaz 2000; OECD 2002; Andersen 2002; Eurostat 2016; Gabrielli et al. 2018; Dijkstra et al. 2019). The idea behind functional areas is the identification of a network of aggregated inbound and outbound movements across spatial structures for a given time scale (for example, daily, intra-weekly, seasonally, etc) according to the scopes of their use.

Thanks to the above mentioned unprecedented collaboration between the European Commission and European Mobile Network Operators during the COVID-19 ongoing pandemic, it has been possible to define the MFAs for 15 European countries (14 member states of the European Union: Austria, Belgium, Bulgaria, Czechia, Denmark, Estonia, Spain, Finland, France, Greece, Croatia, Italy, Sweden, Slovenia, plus Norway), although in this study we will focus on the case of Austria.

Despite the fact that mobility data alone cannot predict future needs (Bwambale et al. 2020), they can show already compelling citizens needs, like transportation or health care facility allocation needs. Moreover, thanks to the capability of collecting mobile data at very high time frequency and space granularity and backward in time, the time evolution of the MFA can indeed show changes or ongoing trends or help to design policies or measure the effectiveness of policies.

This work is organised as follows. “[Functional areas and policy making](#)” section is a brief summary of the different approach to the definition of functional areas and explains the importance of this concept in the field of policy making. “[Mobile positioning data](#)” section describes the essential characteristics of the mobile data used in the analysis. “[Mobility functional areas](#)” section explains in details the method of construction of the MFAs starting from mobility data. As the MFAs are constructed on a daily basis and slightly change daily due to, e.g., intra-weekly human mobility patterns, “[Detecting the persistent MFAs](#)” section describes a denoising method to extract the subset of areas which are consistently connected stable through time. In this respect, the pre-lockdown (or persistent) and lockdown MFAs generated by COVID-19 mobility restriction measures are also discussed. “[Potential applications of mobility functional areas](#)” section presents several potential applications of the concept of Mobility Functional Areas, including transport system, socio-economic analysis, public health, urban planning and environmental risk analysis. To show the effectiveness of the MFAs concept to policy making, “[MFAs and COVID-19 cases during the first wave](#)” section focuses on a public health application. It analyses the impact of the MFAs on the spread of the coronavirus during the first wave of the pandemic in Austria showing that the number of COVID-19 cases increases more within the MFAs than outside. “[Conclusions](#)” section discusses limits and further potential usage of MFAs. The MFAs dataset is openly available to other scholars.

## Functional areas and policy making

As mentioned, the concept of functional areas has a long tradition in settlement geography, urban planning and policy making. These areas have and can be defined in several ways and for different purposes e.g. transportation planning, urban governance and regional development among others. Our proposed concept of highly interconnected geographic zones, the MFAs, are purely data driven were originally developed to fight the COVID-19 pandemic in Europe. Since MFAs are data-driven products generated by natural human mobility, they can be used for several other applications (see “[Potential applications of mobility functional areas](#)” section). A few pre-existing variants of the same concept are the following:

- ‘*Commuting regions*’ the identification of relatively closed regions of daily moves of residing population based on commuting data from censuses (Casado-Díaz 2000; Van der Laan 1998). The aim of the commuting regions is to define geographical areas within which the majority of the work to home travels occur. The definition of commuting regions is particularly relevant for housing and transportation planning purposes.
- ‘*Functional regions*’ a tool used to target areas of specific national and European policies (OECD 2002). There are several natural areas of application of functional regions including employment and transportation policies, environmentally sustainable spatial forms, reforms of administrative regions, strategic level of urban and regional planning and a wide range of geographical analyses (migration, regionalisation, settlement system hierarchisation) (Andersen 2002; Ball 1980; Casado-Díaz 2000; Van der Laan 1998).
- ‘*Functional urban areas*’ cities with their commuting zone (Eurostat 2016; Dijkstra et al. 2019). They are generally identified by a densely inhabited city, together with a less densely populated commuting zone whose labour market is highly integrated with that of the city. They reflect the functional and economic extent of cities and not just their administrative extent. The Functional Urban areas are thus the appropriate spatial level for a more comprehensive urban governance and planning.
- ‘*Overlapping functional regions*’ by Killer and Axhausen (2010). The overlapping functional regions reflect the complexity of commuting flows. One area can be linked to several others and thus could belong to multiple functional areas.

More in general, there exists a long tradition in defining zones related to human mobility. In their extensive review Patterson and Farber (2015) summarize more than 61 different ways of defining Potential Path Areas (PPAs) and Activity Spaces (ASs). The PPAs refer to ‘the spatial extent of where individuals can participate in activities subject to constraints’ (e.g., time, modal availability, etc); ASs refer to locations ‘with which individuals have direct contact as the result of day-to-day activities’. Our MFAs concept, by design, belongs to the ASs strand of research, though MFAs can be potentially used also in the context of PPAs using a probabilistic setup, but this goes beyond the scope of the present study. A further strand of research is that of Transportation Analysis Zones (TAZs), which is generally (though not exclusively) aimed at the optimization of the transport network at suburban level using very high resolution data and geographical/administrative constraints. Unfortunately, the types of aggregated data of this project do not allow this specific application, though high resolution MNO data (XDR or CDR data) may fit the purpose using the same technique applied in this study. As highlighted in Patterson and Farber (2015), most of these methods have in common some types of calculation like, e.g., ellipse methods (Standard Deviation Ellipses, Travel Probability Fields, etc.), Kernel methods (Kernel Density estimation), network analysis, location methods (specific locations at which activities are performed, for example, shopping malls or trip destinations, Points of interests), Minimum Convex-hull Polygons (MCP), etc. Our MFAs rely on the network analysis approach.

Back to the concept of functional areas, the most common data sources for the above-mentioned studies are by far the population censuses and *ad hoc* pilot surveys. In these approaches travel behaviour is captured collecting costly household travel survey data, which are usually very small in terms of sample size, rarely updated and sometimes prone to reporting errors (Bwambale et al. 2020).

Mobile data have been used in the past in some pilot-studies like in Estonia (Novak et al. 2013) or in some regions of Italy in Gabrielli et al. (2018). An attempt to mix survey,

census and mobile data has been proposed as a proof-of-concept by Bwambale et al. (2020) for the case of Bangladesh.

Very recently Facebook released the concept of Commuting Zones<sup>1</sup> (CZs) within their data For Good project. These CZs make use of data coming from the portion of Facebook's user base who opt-in to share their location in the mobile app. The construction of the CZs is also based on a network analysis approach.

The approach used to build the MFAs presented in the next section is special in several ways. Indeed, the approach has been designed to be robust to the diversified data sources available for the 15 countries and also in that it tries to take into account persistent patterns of mobility rather than focusing on a single period as we know that mobility patterns may change due to seasonalities or because there is a strong intra-weekly periodicity.<sup>2</sup> This was possible thanks to the long time series made available through the unprecedented agreement.

As for the other functional areas that make use of mobility data, the MFAs are particularly interesting as the underlying data can be collected at high frequency compared to any survey data, and also covering the whole territory of a country rather than focusing on special cities or routes. Remark also that, since mobile phone services unique subscribers<sup>3</sup> represent about 65% of the population across Europe (GSMA 2020), mobile data can reliably be used to capture the aggregate mobility patterns.

In a policy making framework especially related, but not limited, to the COVID-19 pandemic, the insights resulting from the identification of these new mobility areas may help governments and authorities at various levels:

- (a) to control virus spread by limiting non-essential movements outwards an MFA with a high virus infection rate compared to its neighbour MFAs especially in the initial phase of a virus outbreak;
- (b) to apply targeted and local physical distancing policies in different MFAs, according to their specific epidemiological situation, thus minimizing virus spread and at the same time limiting the economic and social impact of such measures by avoiding not adequate national and universal physical distancing policies;

In the absence of any other information, most of the governments are forced to use administrative areas, such as regions, provinces and municipalities to impose physical distancing measures and mobility restrictions. Nevertheless, administrative boundaries are static and do not reflect actual mobility. On the other hand, both the potential spreading of the virus and the territorial economy strongly depend on local mobility (Iacus et al. 2020).

Although not all of these aspects can be taken into account in this work, the hypothesis is that the implementation of different physical distancing strategies (such as school closures or other human mobility limitations) based on MFA instead of administrative borders might lead to a better balance between the expected positive effect on public health and the negative socio-economic fallout for the country. To this aim, "MFAs and COVID-19 cases during the first wave" section will present a case study of Austria focusing on the impact of MFAs on the spread of the coronavirus.

<sup>1</sup> Details are available here: <https://dataforgood.fb.com/tools/commutingzones/>.

<sup>2</sup> For example, Facebook's CZs are built on the aggregated movements of three months period.

<sup>3</sup> All mobile services subscribers, including IoT, are about 86% of the population, 76% of which real smartphone users.

## Mobile positioning data

The agreement between the European Commission and the MNOs defines the basic characteristics of fully anonymised and aggregate data to be shared with the Commission's Joint Research Centre (JRC.<sup>4</sup>) This section briefly describes the original mobile positioning data from the MNOs; the following section introduces the mobility indicator derived by JRC and used in this research.

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

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

Despite the fact that the ODMs provided by different MNOs have similar structure, they are often very heterogeneous. Their differences can be due to the methodology applied to count the movements, to the spatial granularity or to the time coverage. Nevertheless, each ODM is consistent over time and relative changes are possible to be estimated.

## Mobility functional areas

The construction of the MFAs starts from the ODM at the highest spatial granularity available. This project handled data with different resolution: municipalities, postal code areas, census areas, and different type of grid-level areas. Table 5 in the "Appendix" reports in details the characteristics of the different sets of data used for to calculate the MFAs. The methodology to derive the MFAs is the same for all the 15 countries for which the data are available, though in this study we will focus on Austria only. As it will be explained below, the proposed method is simple with the aim of being robust to the different data specification. Indeed, even for those countries for which the data are made at disposal by more than one MNO, the final MFAs are quite similar. The analysis in this study considers only the mobility from Monday to Friday (with the exclusion of festivities) and this is because weekdays are more relevant for, e.g., transportation system and socio-economic analysis, though in practice the construction of the MFAs can be done for holidays and weekend days with exactly the same steps. The construction of the MFAs starts from origin-destination matrix. Let  $ODM_{d,i,j}$  an element of the ODM matrix for date  $d$  representing the number of movements from area  $i$  to area  $j$ ,  $i, j = 1, \dots, n$

<sup>4</sup> The Joint Research Centre is the European Commission's science and knowledge service. The JRC employs scientists to carry out research in order to provide independent scientific advice and support to EU policy.

$$ODM_{d,i,j}^* = \frac{ODM_{d,i,j}}{\sum_{j=1}^n ODM_{d,i,j}}, \quad i, j, = 1, \dots, n.$$

where  $n$  is the total number of rows and columns of the ODM (which is a  $n \times n$  matrix) and let  $ODM_{d,i,j}^*$  the corresponding element of the ODM normalised by row. Now we transform the  $ODM_d^*$  matrix into a 0/1 proximity matrix  $P_d$  as follows

$$P_{d,i,j} = \begin{cases} 1, & ODM_{d,i,j}^* > \text{threshold}, \\ 0, & \text{otherwise,} \end{cases}$$

where the threshold has been set to 15% and 30% to derive two types of MFAs, that we will call *regular* or *strict* respectively. The threshold of 15% was suggested by Novak et al. (2013) but also used in the functional urban areas concept developed by Eurostat (2016) and Dijkstra et al. (2019). This threshold is a tuning parameter that characterizes the final shapes of the the MFAs. We tested different thresholds above and below 15% including a uniform distribution threshold (i.e., the threshold is set dynamically as the value corresponding to the frequency of a uniform distribution of movements from a given origin to all destinations from this origin that we see in a row of the ODM). Empirically, we can confirm that the 15% seems to be the most effective in isolating stable MFAs for all the countries analysed. It is important to remark that our MFAs are not guided by a theoretical approach as they are completely data driven. Other approaches could have imposed an objective function, like, a minimal number of movements, the maximal length, etc and run an optimization algorithm on top of all the steps presented below to find the *optimal* threshold. So in this study 15% and 30% are informed *a priori* choices. But the main argument on fixing a threshold is that we want to isolate directional movements rather ellipse-like areas. Indeed, the higher the threshold the more the MFAs is concentrated along specific directions. One further remark is that by using relative number of movements, we do not discriminate MFAs based on their absolute volume of movements/population. This is an information that can be checked ex-post.

As the ODM matrix is not symmetric, so is the proximity matrix, which is transformed into an adjacency matrix  $A$  through the following expression:

$$A_d = \frac{1}{2} \cdot (P_d + P_d^T)$$

so that each element of  $A_d$  can take only three values:

- $A_{d,i,j} = 0$  if there are no movements from  $i$  to  $j$  and viceversa (i.e. the two areas are not connected);
- $A_{d,i,j} = 0.5$  if there are movements only in one direction, either  $i$  to  $j$  or  $j$  to  $i$ .
- $A_{d,i,j} = 1$  if there are movements in both directions, from  $i$  to  $j$  and from  $j$  to  $i$ ;

From the adjacency matrix we construct a directed<sup>5</sup> graph where the vertex represent the areas  $i = 1, \dots, n$  and the edges are weighted according to the matrix  $A$ . The MFAs are calculated using a community detection technique called *walktrap* algorithm (Pons and Latapy 2006), which finds communities through a series of short random walks. This approach is

<sup>5</sup> An undirected graph could be used as well, but we use a directed graph in view of the community detection algorithm used later on.

different from the *intramax* algorithm used in Killer and Axhausen (2010); Novak et al. (2013). The idea is that these random walks tend to stay within the same community. The goal of the walktrap algorithm is indeed to identify the partition of a graph that maximises its modularity. The modularity of a graph is the index designed to measure the strength of division of a network into modules (also called groups, clusters or communities). This is why the walktrap algorithm is well suited for finding clusters of fully interconnected cells where most of the movements are internal.

All the communities with only one member, i.e. those without inbound and outbound movements over the 15% (or 30%) threshold, are collapsed into a single big fictitious area representing the territory that either cannot be identified as a pure MFA, or it is just a collection of atomic (mobility-wise) areas. As this algorithm is data-driven, the number of MFAs is not prescribed in advance but is the outcome of the algorithm itself. For example, for Austria we have more than 200 MFAs per day, some of them are very small, others are quite large. In all events, their shapes depend only on the mobility itself.

### Mobility patterns and mobility functional areas

It is well known and expected that mobility changes between weekdays, weekends and holidays, but there might be also an internal variability within the working week as well as across weeks (e.g., not all Mondays are *exactly* the same in terms of mobility). Therefore, the MFAs might be slightly different from day to day. In order to denoise the MFAs we will first measure how much the MFAs are stable and consistent through time through a similarity index.

We make use of the following similarity index (Gravilov et al. 2000) between two sets of groups of labelled  $G = \{G_1, \dots, G_K\}$  and  $G' = \{G'_1, \dots, G'_{K'}\}$ , where  $K$  and  $K'$  are not necessarily equal. For example, when two cluster algorithms are applied to the same set of  $n$  observations, it might happen that algorithm 1 produces  $K$  groups and algorithm 2 produces  $K'$  groups. In this cluster analysis notation, each  $G_i$  (or  $G'_j$ ) is a subset of indexes corresponding to observations that fall in group  $i$  (respectively  $j$ ). In our case,  $G_i$  is the subset of areas id's corresponding to the  $i$ th MFA. The similarity index is defined as

$$\text{Sim}(G, G') = \frac{1}{K} \sum_{i=1}^k \left\{ \max_{j=1, \dots, K'} \text{sim}(G_i, G'_j) \right\} \tag{1}$$

where

$$\text{sim}(G_i, G'_j) = 2 \frac{|G_i \cap G'_j|}{|G_i| + |G'_j|} \in [0, 1], \quad i = 1, \dots, K, j = 1, \dots, K'.$$

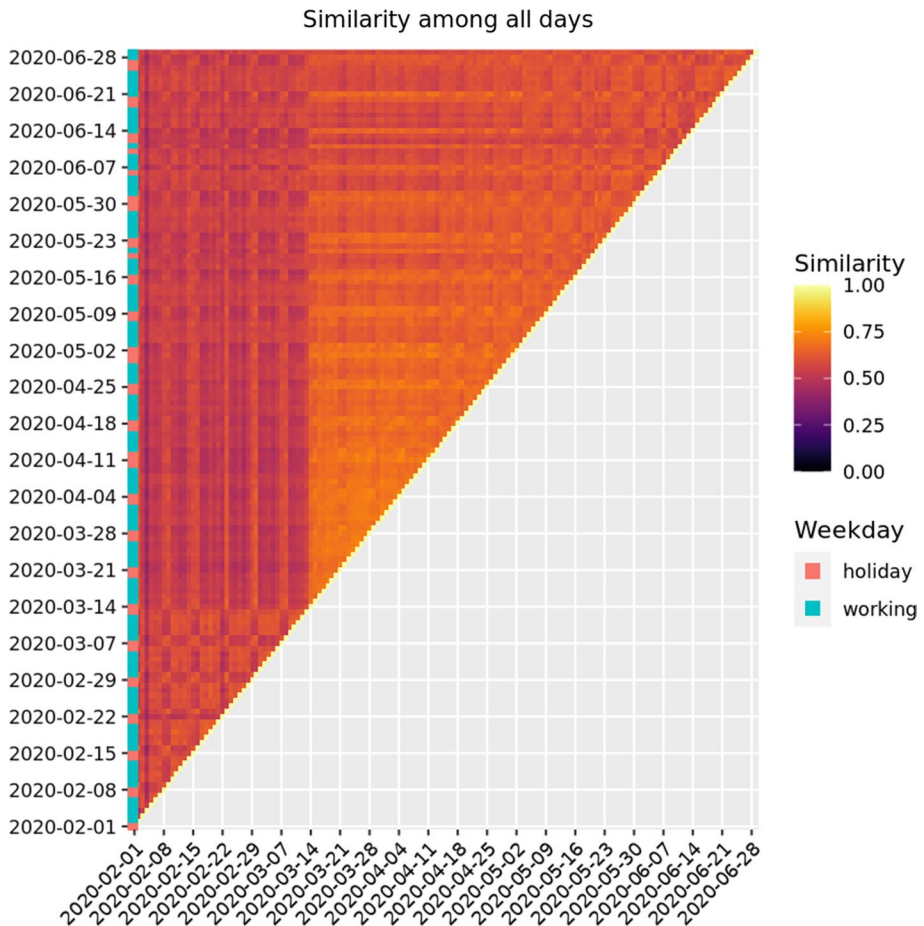
with  $|B|$  the number of elements in set  $B$ .

The similarity index is such that  $\text{Sim}(G, G') \in [0, 1]$  but it is not symmetric, therefore in order to have a symmetric measure we consider

$$\overline{\text{Sim}}(G, G') = \frac{1}{2} (\text{Sim}(G, G') + \text{Sim}(G', G')).$$

Figure 1 is a heatmap representation of the matrix of similarity index among all MFAs for the period 1 February 2020 - 29 June 2020. Darkest-bluish zones are most different, whereas lighter-reddish are the most similar. It is interesting to observe the difference



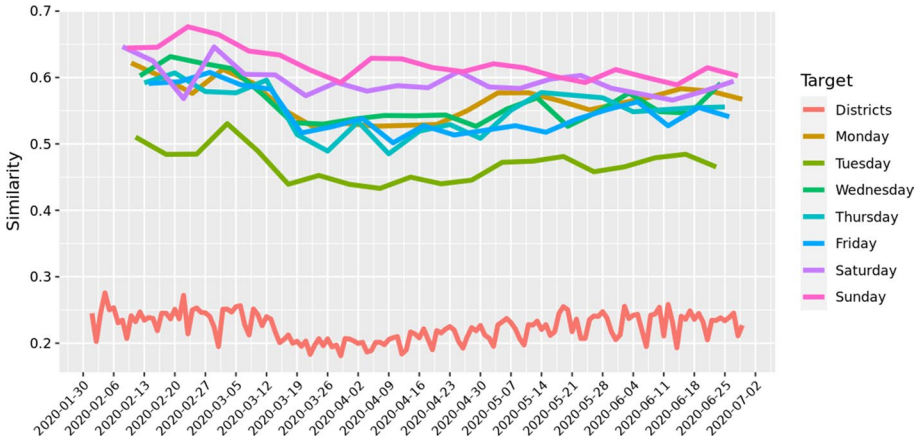


**Fig. 1** Similarity index matrix of all MFAs in Austria. Period: 1 February 2020–29 June 2020. It is clear that the persistent (pre-lockdown) MFAs are different from the lockdown MFAs and that slowly, the mobility is going back to normality by the end of June 2020. Lockdown was enforced in Tyrol on 13 March 2020 and nationwide on the 16th of March 2020

between before and after the lockdown (14 March 2020), then a slow recovery to normality. It is also worth noticing that holidays and weekends have clearly different mobility patterns than weekdays and that these MFAs are different from the administrative borders (Austrian districts), especially during weekdays.

To put in evidence the intra weekly patterns in Fig. 2 we calculate the similarity of the MFAs for the same day of the week. We start from the first Monday of the data and we calculate the similarity of it with all the subsequent Mondays. Similarly for the Tuesdays, etc. We also calculate similarity of the daily MFAs with the administrative borders. What clearly emerges from Fig. 2 is that each day of the week has an almost stable pattern before and after the lockdown and in this sense the MFAs are similar among them for the same day of the week. Further, we can notice that the similarity of the MFAs with the administrative borders is quite low, meaning that indeed MFAs do not coincide with the latter. The effect of lockdown is also seen in this curve.





**Fig. 2** Intra weekly similarity (Mondays vs. first Mondays, etc) of daily MFAs and with respect to the Austrian districts (daily MFAs with respect to administrative borders). Lockdown was enforced in Tyrol on 13 March 2020 and nationwide on the 16th of March 2020

### Detecting the persistent MFAs

As seen in the previous section, the MFAs have daily patterns, they change between before and after the lockdown is in force and tend to go back to their original shapes after the ease of containment measures. Moreover, MFAs shows time-variability also for the same week-day; thus, in order to fully exploit their potential, a stable version of the MFAs needs to be identified. Since the number of MFAs changes day by day and the same area may move from an MFA to another (changing the MFA label associated to it), we apply a *CO*-association method. The *CO*-association method (*CO*) avoids the label correspondence problem. It does so by mapping the ensemble members onto a new representation where the similarity matrix is calculated between a pair of objects in terms of how many times a particular pair is clustered together in all ensemble members (Fred and Jain 2005). In other words, *CO* calculates the percentage of agreement between ensemble members in which a given pair of objects is placed in the same MFA.

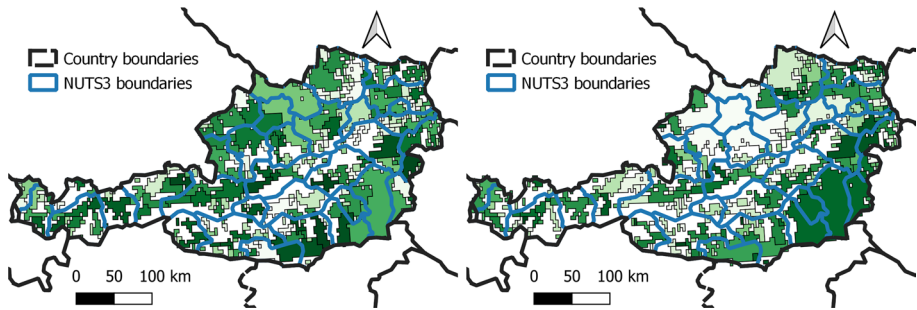
As the first lockdown in Austria was enforced on 16 March 2020, we focus on the weekdays from 1 February 2020 to 15 March 2020.

Let  $d$  be the data of one of these  $D$  weekdays and  $MFA_d$  the set of mobility functional areas obtained on day  $d$ . We then evaluate the *co-association matrix*

$$CO(x_i, x_j) = \frac{1}{D} \sum_{d=1}^D \delta(MFA_d(x_i), MFA_d(x_j))$$

where  $X_i$  and  $x_j$  are the areas and  $MFA_d$  is the set of MFAs for day  $d = 1, \dots, D$  and  $\delta(\cdot, \cdot)$  is defined as follows:

$$\delta(u, v) = \begin{cases} 1, & \text{if } u \text{ and } v \text{ belong to the same MFA,} \\ 0, & \text{otherwise.} \end{cases}$$



**Fig. 3** Persistent and lockdown MFAs for Austria compared to NUTS3 districts (blue lines). Whereas before the lockdown the MFAs extend across provinces, after the lockdown their area generally reduces and they mostly lay within districts' borders (with some exceptions) or they completely disappear as all mobility remains within individual areas. Areas where connectivity is below 15% (no apparent stable direction is observed in the mobility flows) are transparent. The color scale is only meant to distinguish the MFAs. Colors of the left and right maps are different as the number of MFAs is different. (Color figure online)

Then, as our scope is to obtain a *persistent* version of the MFAs, we further threshold the  $CO$  matrix so that all entries below 50% are set to 0 and those higher or equal 50% are set to 1 (it means that only cells falling in the same MFA at least 50% of the times are associated with that MFA) leading to a new matrix  $\overline{CO}$ .

Then again a directed graph is built with this matrix using the entries of the  $\overline{CO}$  matrix to weight the edges and applying the walktrap algorithm to obtain the final persistent MFA. The same procedure is replicated for the lockdown dates, ending up with a different set of stable MFAs that we denote by lockdown-MFA. With these two sets of MFAs at hand, we further test if they are meaningful to the analysis. It turns out that these stable MFA are in fact reasonably well defined.

We then apply the symmetric similarity index  $\overline{\text{Sim}}(\cdot, \cdot)$  for all the daily MFAs against the persistent MFAs, the lockdown MFA and the districts (NUTS3). Figure 3.

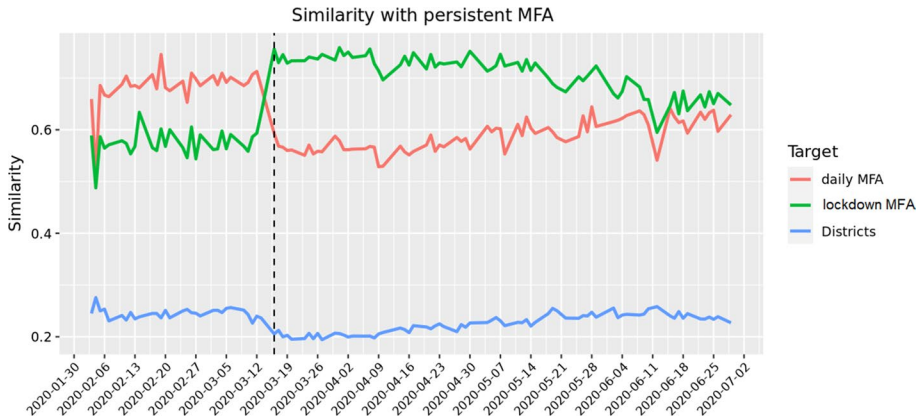
In Fig. 4 we calculate the similarity index between the daily MFAs against the lockdown MFAs and, against the districts. It quite evident the abrupt change in the shape of the MFAs on March 16h which corresponds to the stay-at-home nation wide order imposed by the Austrian authorities.

## Potential applications of mobility functional areas

The MFAs, originally developed in the framework of the non-medical interventions to the fight the spread of COVID-19 in Europe, have many other fields of application. In this section, we briefly take an overview of some of these fields.

### Transport system

As said, the MFAs are data-driven products generated by natural human mobility, which is in fact strictly related to available local transport system; this is indeed a bi-directional relationship. At the same time, the MFAs can inform about possible new directions in which the transport system can be enhanced. Another interesting aspect of the MFAs is that, especially around big cities, they are not necessarily made of contiguous zones; although this



**Fig. 4** Similarity of daily MFAs (orange line), lockdown MFAs (green line) and Districts (blue line) with respect to persistent MFAs. Daily MFAs are very similar to the persistent MFAs before lockdown which was enforced on 16 March 2020, whereas they are more similar to lockdown MFAs after. A return to normality is slowly appearing. Once again, the MFAs are quite different from administrative borders. (Color figure online)

is an artifact, or better a feature of the MFAs due to how they have been designed, starting from Origin-Destination Matrices (ODMs). This does not allow showing that when going from **a** to **c** one can stop by a bar to take a coffee in **b**. For example, in Fig. 5 we can see some typical stylized facts of MFAs:

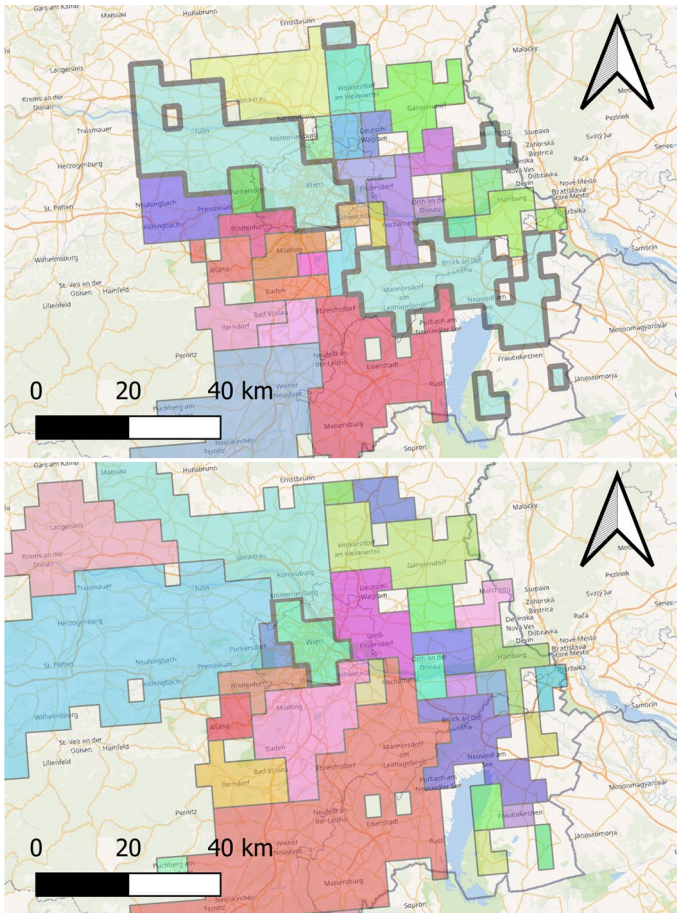
- MFAs in some cases follow already existing infrastructures, like highways;
- MFAs are not contiguous;
- MFAs are not all of the same size and this is clearly related to the relative importance of the geographical zones involved.

The largest MFA in the top map of Fig. 5 is clearly made of contiguous and not-contiguous areas and its development follows the the most important highway crossing the capital. These zones, which are highly interconnected by definition of MFA, represent important highlights to eventually improve the transport system accordingly. The bottom map in Fig. 5 shows how the MFAs have shrunk during the lockdown.

Following the same approach used for the Vienna example, but outside the pandemic context, MFAs calculated over two different time periods and compared with each other can also help monitoring the actual usage of transport infrastructures in order to support their management. Moreover, by analysing daily MFAs, local administrations can monitor how mobility changes in occasion of major mass-gathering events, learning how to plan local mobility restrictions and alternative paths.

**Socio-economic impact of the pandemic**

By comparing the two maps in Fig. 5, it is evident how MFAs can show major changes in human mobility patterns. Eventually, with MFAs built starting from higher spatial-resolution data, it would have been also possible to asses the socio-economic impact of the lockdown measures at urban level. As an example, one would have been be able to identify



**Fig. 5** The persistent MFAs around Vienna before (top) and during the lockdown (bottom). Colors are randomly assigned to the MFAs, without any relation between top and bottom maps. The largest persistent MFA around Vienna (top map), marked with a bold gray border, is made of non contiguous zones and follows the main highways passing through the capital. This MFA is highly shrunk during lockdown (bottom map). On the other hand, some of the MFAs merge during lockdown and there are no MFAs made by non contiguous zones. This seems indicating that movements are likely to be more local during the lockdown. Background map Open Street Map: Road network in orange; Country boundaries in grey; water surfaces in light blue. (Color figure online)

neighborhoods where the residential population kept moving for working-related reasons (because they could not telework or they were essential workers), despite the mobility restrictions and the risk of being infected. Another example is the ability to monitor whether or not key commercial areas (such as malls, shopping hubs, etc.) had remained connected in terms of mobility with the high-population-density neighborhoods.

## Public health

The MFAs were originally designed to support the fight to the spread of COVID-19 in Europe. Their role in the public health sector, specifically during a pandemic, is therefore evident and is further developed in “[MFAs and COVID-19 cases during the first wave](#)” section. Beside their employment in an emergency context, MFAs could be also exploited to plan and monitor the accessibility of local health units, main hospitals and any other type of health facility.

## Urban planning

The persistent MFAs can help authorities and spatial planning agencies to take more evidence-based and informed decisions. Sustainable spatial planning and spatial management requires the identification of the optimal location for both private and public facilities and services that minimize transportation needs while maximizing service availability. For example, MFAs can assist spatial planners and developers on identifying the optimal location of new housing developments, schools and sports facilities.

## Environmental risk and pollution

Since the MFAs show human mobility, when it comes to environmental monitoring and risk management, they can be profitably used to design the displacement of monitoring stations, in order to measure those pollutant associated to road-mobility. Analogously, MFAs could support the planning of new waste-to-energy plants, wastewater treatment plants biomass plants, etc, outside the main human-mobility areas.

## Demography and migration

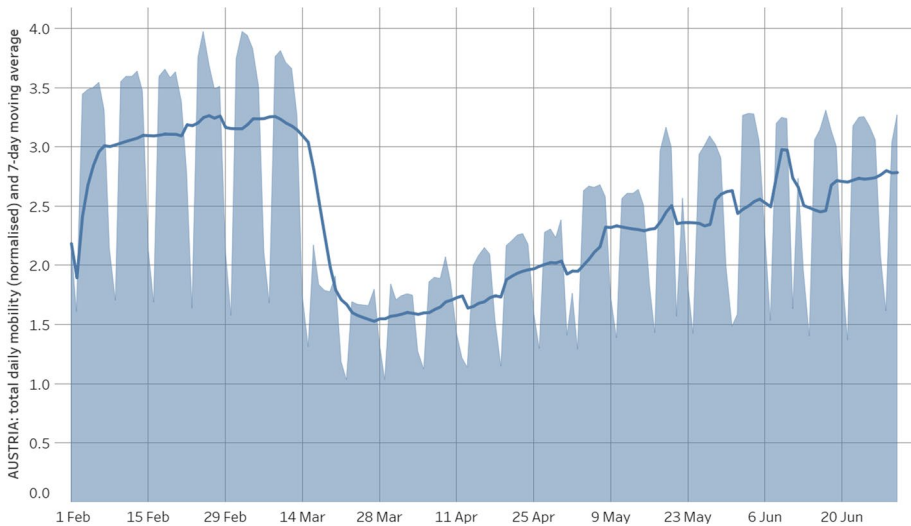
MFAs can be used to identify disadvantaged/isolated areas which are not well connected to economically developed MFAs. These areas are more likely to experience an aging population as well as a depopulation. Besides, without any administrative intervention, these areas are likely to be less attractive to international migrants (Aurambout et al. 2021).

## MFAs and COVID-19 cases during the first wave

In this section, we analyse the role of MFAs in the contest of the spreading of COVID-19. The objective is testing the hypothesis that MFAs are correlated to the evolution of the number of cases until strict measures to limit mobility are enforced. For this reason, we focus on the period between 15 March 2020 and 30 April 2020, monitoring the number of cases in the areas laying within or outside the persistent MFAs. It is worth noticing that the actual timeline of the measures taken is the following:<sup>6</sup> (a) On 13 March 2020 the entire valley surrounding Ischgl was in lockdown; (b) On 15 March the state of Tyrol was in lockdown; (c) From 16 March a nationwide lockdown was implemented in Austria. Thus given

---

<sup>6</sup> Source Wikipedia: [https://en.wikipedia.org/wiki/COVID-19\\_pandemic\\_in\\_Austria](https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Austria).



**Fig. 6** Relative mobility change in Austria during the period 1 February 2020–30 June 2020

**Table 1** The MFAs with more than 20000 inhabitants

MFA	Population	km <sup>2</sup>	Districts that intersect the MFA
6	22,126	96	Mödling, Sankt Pölten (Land), Wien (Stadt)
137	27,870	64	Korneuburg, Wien (Stadt)
177	29,479	32	Bruck an der Leitha
133	47,442	80	Gänserndorf, Wien (Stadt)
132	59,897	48	Wien (Stadt)

See also Fig. 7

**Table 2** Distribution of the areas of the MFAs in km<sup>2</sup> before (persistent) and during the lockdown

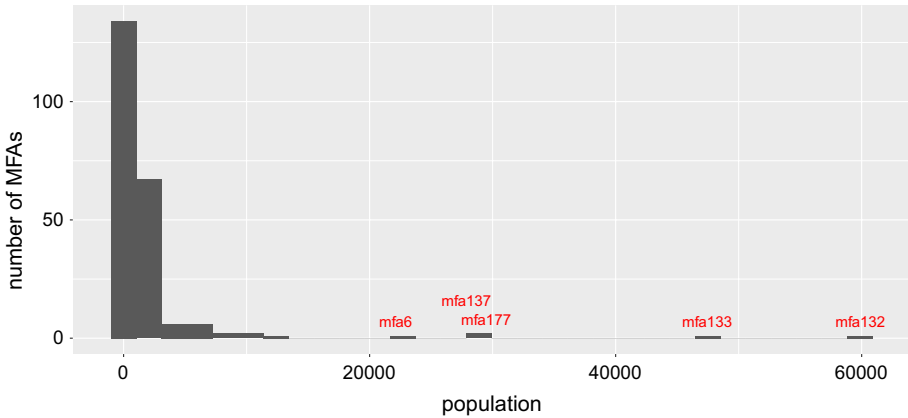
MFA type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Persistent	16	48	89	248	246	4765
Lockdown	31	33	65	159	143	2302

the limited period of time that local measures were in place we do not expect any endogeneity issues in the statistical analysis that follows. Figure 6 shows the relative change in mobility for Austria during the period 1 February 2020–30 June 2020.

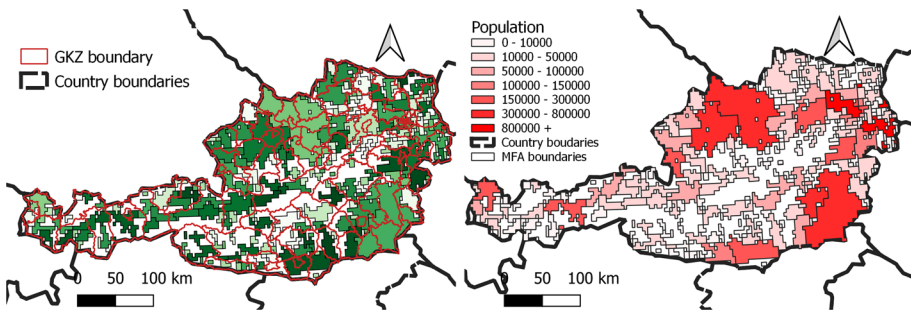
It is known in the literature that it takes about 14 days (Iacus et al. 2020) to see the decrease in the number of cases due to a lockdown, so we should expect a peak around the 30th of March and then a decrease of cases as indeed Fig. 11 shows.

In Austria we have 222 persistent MFAs (plus the residual territory) built on 4789 areas identified by the MNO, while the number of political districts (GKZ) is 94. The GKZ do not match with the NUTS3 districts used in “[Mobility functional areas](#)” and “[Detecting the persistent MFAs](#)” sections and these are the geographical units for which





**Fig. 7** Distribution of population by MFA. Only 5 MFAs are large, i.e. more than 20,000 inhabitants, in population size and all but one overlaps with Wien district (see also Table 1)



**Fig. 8** MFAs compared to GKZs borders (left) and the estimated population on the MFAs (right). The MFAs also cross the administrative borders on which the health data are available

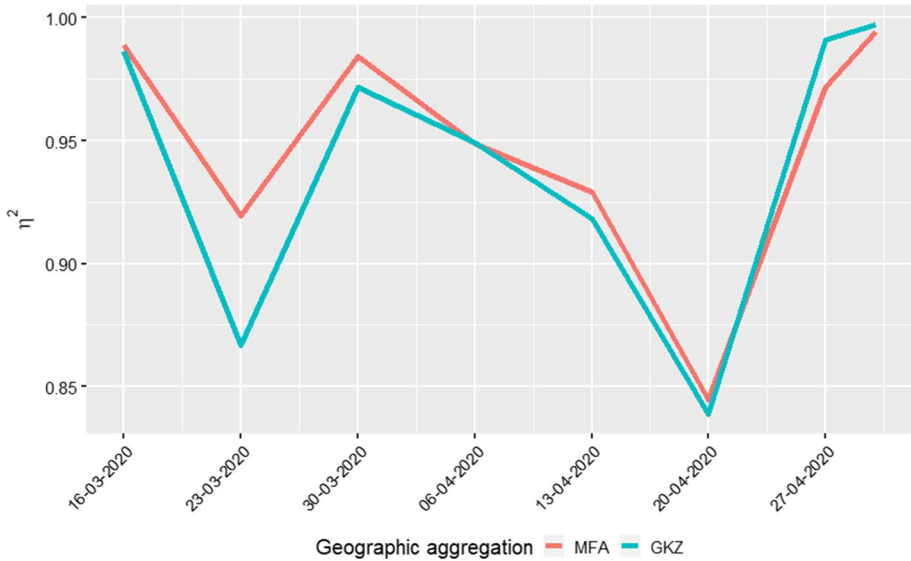
epidemiological data is available. Of the 4789 zones, 3409 belong to a proper MFA and 1380 are in the residual territory of Austria. Not all the MFAs are big enough to be relevant either because their population is small or because their territorial extent is very limited. Figure 7 shows the distribution of MFAs by population size and Table 1 shows the corresponding intersecting districts. As it can be seen, in all but one case, these MFAs contain the territory of Wien. Table 2 reports the extension of the MFAs in km<sup>2</sup>. Using high resolution Facebook population data,<sup>7</sup> we reconstruct the population of each of 4789 the zones. Figure 8 shows the shapes of the MFAs compared to those of the political districts for which health data are available and also the estimated population for each zone of the MFAs.

It is also worth to mention that the daily density of movements<sup>8</sup> in the *residual* MFA is equal to 0.00012, while on average the other MFAs have a density of 15.016 (Q1 =

<sup>7</sup> <https://dataforgood.fb.com/tools/population-density-maps/>.

<sup>8</sup> By daily density of movements we mean the average number of movements per day, in the pre-lockdown period, per cell of the grid that defines each MFA.





**Fig. 9** The index  $\eta^2_{Y|G}$  by calendar week by group  $G = MFA$  or  $G = GKZ$ . The higher the value of  $\eta^2_{Y|G}$ , the more  $Y$  is determined by  $G$ . Notice that  $\eta^2_{Y|MFA} \geq \eta^2_{Y|GKZ}$  till 20 April 2020, then the reverse is true. This is expected because few weeks after the national lockdown, the impact of the MFA does no longer appear as the mobility has been reduced or stopped completely

4.065, median = 8.100, Q3 = 13.680, max = 294.111), meaning that most of the mobility is observed in the identified MFAs. Notice further that the residual MFA is made of 1553 cells of the grid, while on average an MFA is made of 14 cells with a maximum of 279 (Q1=3, Q3=15).

As said, the COVID-19 health data for Austria are available only at political district level from the open data repository<sup>9</sup> of the Federal Ministry of social Affairs, Health, Care and Consumer Protection.<sup>10</sup> As it is known that the number of infection cases is proportional to the population size, we redistribute the number of COVID-19 cases on the 4789 zones proportionally to the population size of the zones within each district. Figure 10 shows the number of cases in the last seven days for four equally spaced dates from 15 March to 30 April 2020. These maps seem to confirm qualitatively the role of the MFAs in the spread of the virus. To assess this qualitative evidence with a quantitative index, prior to the statistical analysis that will follow, we can use the index of determination (Pearson’s correlation ratio)  $\eta^2_{Y|G}$  (Fisher 1926) which measures the dependency of a continuous variable  $Y$  in terms of a categorical variable, usually a grouping variable. The index is based on the decomposition of the total variance of  $Y$  into the sum of the variance *within* the groups ( $\sigma^2_W$ ) and the variance *between* the groups ( $\sigma^2_B$ ):

$$\sigma^2 = \sigma^2_W + \sigma^2_B = \frac{1}{N} \sum_{j=1}^K n_j \sigma_j^2 + \frac{1}{N} \sum_{j=1}^K (\bar{x}_j - \bar{x}_N)^2 n_k$$

<sup>9</sup> <https://github.com/statistikat/coronaDAT>.

<sup>10</sup> <https://info.gesundheitsministerium.at/>.

**Table 3** The index  $\eta_{Y|G}^2$  by calendar week by group  $G =$  MFA or  $G =$  GKZ

MFA	GKZ	Date range	Clendar week
0.99	0.99	2020-03-15–2020-03-16	11
0.92	0.87	2020-03-17–2020-03-23	12
0.98	0.97	2020-03-24–2020-03-30	13
0.95	0.95	2020-03-31–2020-04-06	14
0.93	0.92	2020-04-07–2020-04-13	15
0.84	0.84	2020-04-14–2020-04-20	16
0.97	0.99	2020-04-21–2020-04-27	17
0.99	1.00	2020-04-28–2020-04-30	18

The higher the value of  $\eta_{Y|G}^2$ , the more  $Y$  is determined by  $G$

where  $\bar{x}_j$  and  $\sigma_j^2$  are respectively the mean and the variance of group  $j$ ,  $\bar{x}_N$  is the global mean, and  $N = n_1 + n_2 + \dots + n_k$ , is the total number of observations. The index of determination is given by the formula

$$\eta_{Y|G}^2 = \frac{\sigma_B^2}{\sigma^2} \in [0, 1]. \tag{2}$$

When the within variance is small it means that the grouping variable creates homogeneous groups in terms of the variability of  $Y$ , in which case  $\eta_{Y|G}^2$  reaches an high value. To compare the effectiveness of the MFAs to capture the dynamics of the pandemic, we calculate the  $\eta_{Y|G}^2$  index using the MFAs or GKZs as grouping variables, and we take the sum of number of cases in the past seven days in each MFA or GKZ as response variable  $Y$ . The results are given in Table 3 and plotted in Fig. 9. The results show that most of the time the value of  $\eta_{Y|MFA}^2 \geq \eta_{Y|GKZ}^2$  from 15 March 2020 till 20 April 2020, then the reverse is true. This is expected and confirmed by later analysis, because the impact of the MFA can be present only up to few weeks after national lockdown are put in place.

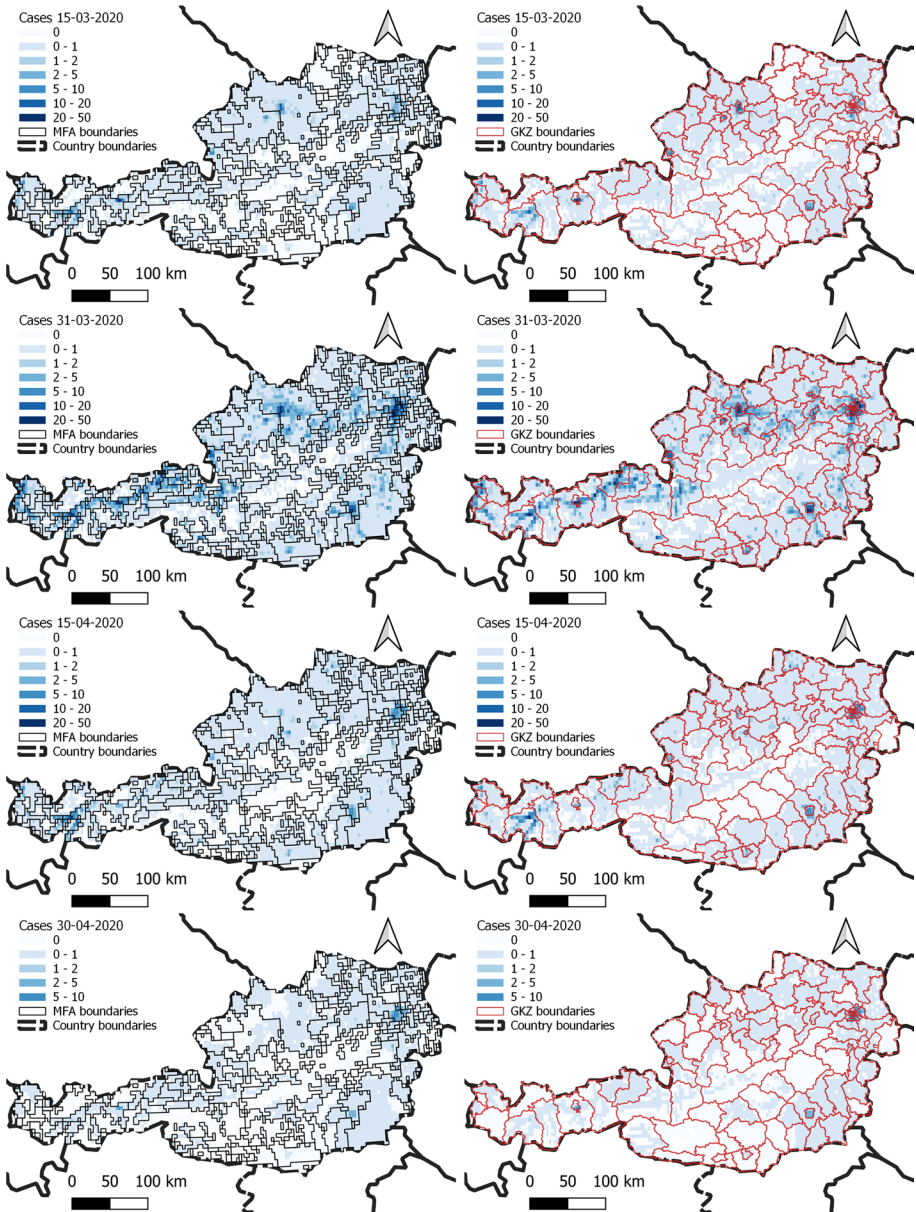
We now test our hypothesis in two ways. First of all we consider a simple linear model for the log number of cases in the last 7 days and the indicator function *mfaInd* which takes value equal to 1 if the zone belongs to a MFA and zero otherwise. We also control for the *population* size of the zone. The model is as follows

$$y_{i,t} = \alpha + \beta \cdot \text{mfaInd}_i + \gamma \cdot \text{population}_i, \tag{3}$$

for  $t = 15$  March 2020, ..., 30 April 2020,  $y_{i,t}$  is the logarithm of the number of COVID-19 cases in the last 7 days for the zone  $i, i = 1, \dots, 4789$ .

Figure 11 shows the sign of the standardized<sup>11</sup> coefficient of the *mfaInd* as well as those of *population* and the constant of the model, along with the corresponding adjusted  $R^2$  goodness of fit index for the model. Thanks to the standardization, we can see that, although the population size has the main effect, the belonging to an MFAs is also always significant (the plots shows only coefficients whose  $p$ -values are less than 0.01). We use

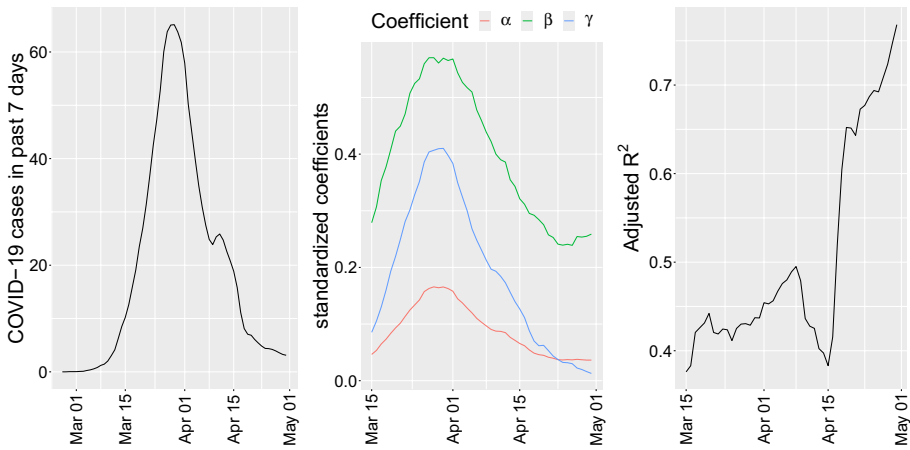
<sup>11</sup> As *mfaInd* is a categorical variable, to make sense of the standardization we apply Gelman’s rule to standardize the *population* by two standard deviations. See <http://www.stat.columbia.edu/~gelman/research/unpublished/standardizing.pdf>.



**Fig. 10** The number of cases in past 7 days for the dates 15 and 31 March, 15 and 30 April 2020 on the MFAs (left) and on the GKZs (right). It looks like the COVID-19 spread seems to follow more closely the MFAs borders than the GKZ borders

the number of total cases in the past 7 days in order to take into account lagged effects of mobility on the evolution of the pandemic.

It turns out that the MFA indicator function has always a positive coefficient  $\beta$  with  $p$ -value less than 0.001, i.e., always statistically significant. Moreover, the overall fit of the



**Fig. 11** Left to right: the number of cases in past 7 days, the value of the coefficients  $\alpha$ ,  $\beta$  and  $\gamma$  in model (3) and the adjusted  $R^2$  of the estimated model (3) for each date. Coefficients are always significant at level 0.01

model is acceptable being in the range 0.4-0.5 most of the times during the peak of the epidemic. The fact that towards the end of the period the MFA indicator variable has a coefficient decreasing to zero is due to the fact that the mobility pattern changed due to national lockdown measures (see also Fig. 6) and thus the MFA are quite different.

Model (3) is capturing an average effect of MFA impact on the number of COVID-19 cases. To disentangle the impact of each MFA on the pandemic, we use a random effect linear model, also known as mixed effect model (Bates and DebRoy 2004) or hierarchical model (Gelman and Hill 2006). In the form we use it is just a regression with clustered errors that depend on a grouping variable. In our case, the dependent variable is again  $y_{i,t}$  but we replace in (3) the indicator function  $mfaInd$  with the MFA categorical variable to obtain Eq. (4) below. The MFA variable has 223 different values, so it implies 222 new dummy variables in the model, most of them correlated, and for which there are not much observations per groups as some MFA are made of two or few more zones. To solve this problem we apply the random effect model as implemented in Bates et al. (2015). In practice, we cluster the observations in groups identified by the political districts (GKZ). Exploiting the correlation between the GKZs and the MFAs it is possible to estimate all the variances of the MFA dummy variables and hence test the significance of each coefficient. The random effect model can be written in this form, where “| GKZ” means *conditionally to* in the sense of Bates et al. (2015):

$$y_{i,t} = \alpha + \beta \cdot mfa_i + \gamma \cdot population_i | GKZ, \tag{4}$$

Figure 12 shows the estimated coefficients, with their respective confidence intervals, for the same four dates of Fig. 10. Only the coefficients which are significant at 0.01 are shown. It can be noticed a couple of things: (1) the number of MFAs retained changes through time though those MFAs that remain have a positive coefficient, meaning that

**Fig. 12** Left to right: coefficients of the MFA categorical variable for different dates. Only the coefficients which are statistically different from zero are shown. The only notable MFA is the MFA number 181, for which the sign of the coefficient is negative. The colors of the points are not relevant. (Color figure online)

being in those particular MFAs increases the number of cases. (2) it is possible to see that the only MFA which has a negative coefficient is the number 181 but this happens at the end of the period of analysis.

A dynamical view which involves all significant MFAs during the whole period of the analysis is presented in Fig. 13 which shows the heatmap of the corresponding MFAs coefficients which are statistically significant at 0.01 level in all the dates analyzed. It is again clear that the only special MFA is the number 181 that becomes significant and negative starting from 17 April 2020 when all the other MFAs are no longer significant. But around 15 April, the number of cases is also decreasing very quickly towards zero, meaning that there is essentially no signal in the data, i.e., there are no longer cases to analyze.

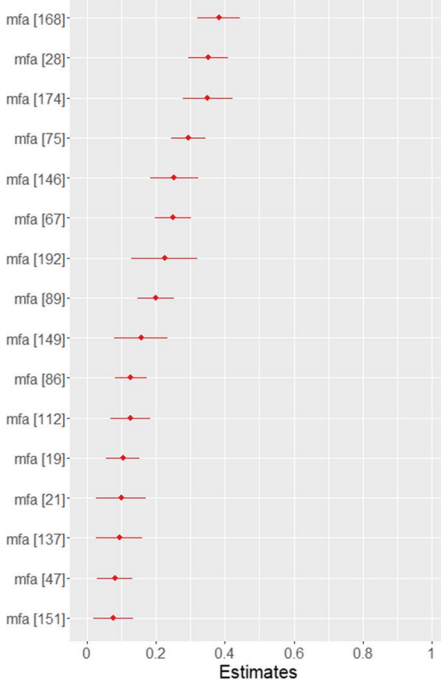
This MFA number 181 is made of zones in the Wien Stad and Sankt Pölten Land, a small district on the west border to the city of Vienna. The conclusion here seems to be that when the number of cases goes down due to countermeasures (lockdowns, curfews, etc) the impact of the persistent (pre-lockdown) MFAs is clearly vanishing, while until the lockdown, the MFAs are largely driving the pandemic. The negative impact related to the MFA number 181 is mainly due to the lack of cases for all the other MFAs and a few left cases in that MFAs, which naturally leads to a negative sign of the regression coefficient. The other MFA which shows a negative coefficient in a single date, is the MFA number 219 that corresponds to the small village of Weiz. As this is one date only, we do not think that this is a statistical evidence so we do not comment any further.

Table 4 shows that in terms of population size and therefore number of COVID-19 cases, the MFAs which are not statistically significant are in general smaller. This type of results is quite natural as it means that the areas to be contained or monitored are correctly identified in our approach. This implies also that the MFA with higher population can be isolated if (1) they have too many cases, in which case there is high probability of exporting cases; or if (2) there are no infections, in which case it is possible to avoid the import of new cases (Iacus et al. 2020).

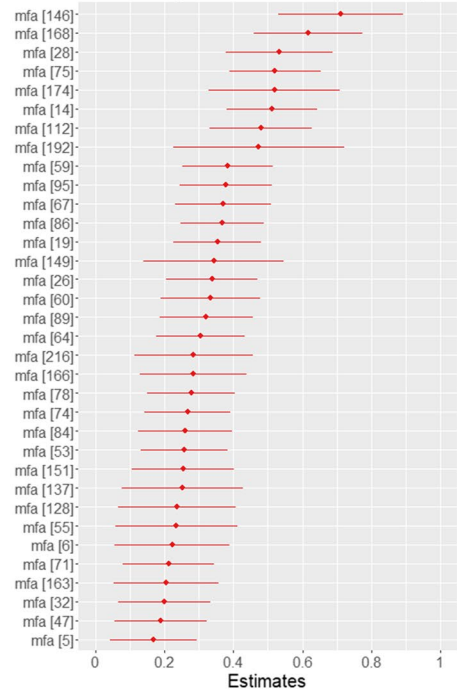
To summarize, this analysis clearly shows the impact of MFAs on COVID-19 spread (along with other causes) and can inform policies in future waves. Indeed, a potential implication of this analysis is that, if new cases arises in a zone belonging to a specific MFA, that MFA should be isolated (or tested extensively) to efficiently contain further spread of the virus, rather than focusing on the whole or single district (when MFAs crosses districts) or instead of blocking the whole country, in the spirit of the zero-covid strategy.<sup>12</sup>

<sup>12</sup> <https://www.euronews.com/2021/02/23/what-is-a-zero-covid-strategy-and-could-it-be-implemented-in-europe>.

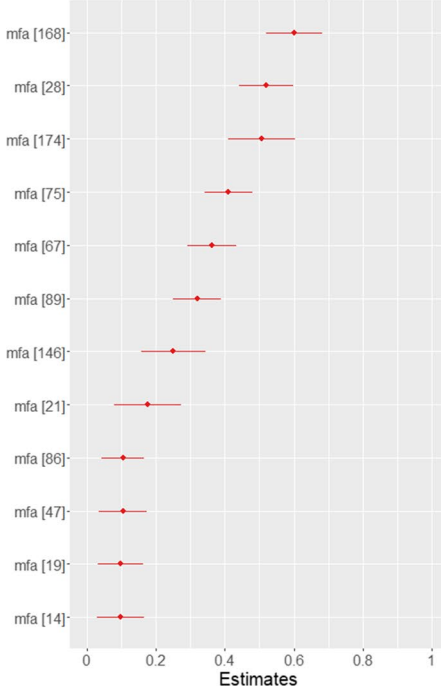
Cases7Days : 2020-03-15 (thr=0.01)



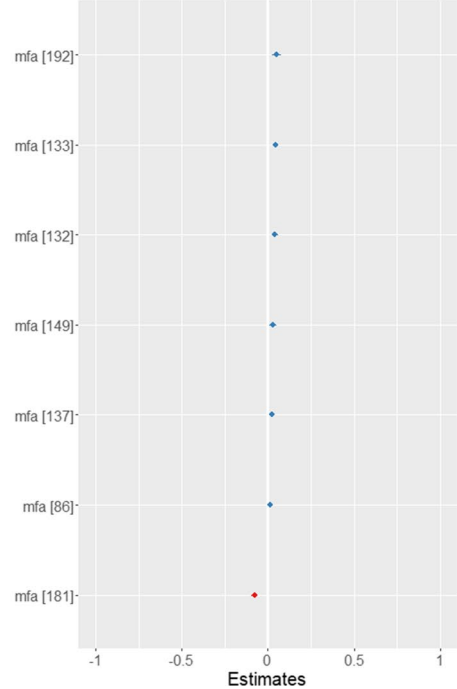
Cases7Days : 2020-03-31 (thr=0.01)



Cases7Days : 2020-04-15 (thr=0.01)



Cases7Days : 2020-04-30 (thr=0.01)



## Conclusions

The present work, in line with the literature on functional regions, puts in evidence how administrative borders typically differ from actual commuting patterns, which are shaped by human mobility. An innovative way to reflect human mobility patterns is given by the ‘*Mobility Functional Areas*’ (MFAs). The MFAs are data-driven geographic entities derived from fully anonymised and aggregated data originally provided to the European Commission by several European Mobile Network Operators (MNOs) in the framework of the non-medical interventions to contrast the spreading of COVID-19 pandemic in Europe.

Though slightly changing day by day, MFAs are essentially persistent in time and identify clear intra-weekly patterns. We have shown how this patterns have changed in Austria following mobility-restrictions imposed by the national authorities to limit the spreading of the virus. Analogous changes in mobility patterns have been measured, with different extents, in all the European countries that we have analysed. In fact, national lockdowns and similar mobility-restriction measures at the local level have not only reduced the volume of mobility but have also drastically changed the mobility patterns and this “shrinking effect” is clearly reflected in the shape and distribution of the MFAs.

The paper shows how, at a very local scale, MFAs can support a more efficient planning of the transport system, as well as its continuous monitoring. It also introduces several other potential fields of application for the MFAs (i.e. socio-economic, public health, urban planning, environmental risk and pollution, and demography and migration), which should be further investigated in future researches. Finally, it provides statistical evidences that during the initial phase of the pandemic, the MFAs played an important role in the spread of the virus. On this bases, an enforcement of mobility restrictions based on MFAs rather than administrative areas would have been more effective in limiting both the spreading of the virus and the socio-economic impact of the restrictions. That said, the authors acknowledge that from an operative point of view, the enforcement of mobility restrictions based on administrative borders (which are well recognised by citizens and reflect administrative and law-enforcement areas) would be much easier to implement than one based on MFAs, which very often cross these administrative borders. For example, in case of a local outbreak in a municipality, instead of applying restrictive measures in a whole region or country health authorities can only apply them in the municipalities that belong to the same MFA. For this reason, coordination efforts between local and regional authorities are required.

While this study focuses on Austria, the full study comprises the analysis of “**Mobility functional areas**” to “**Potential applications of mobility functional areas**” sections, for 15 European countries as shown in Table 5 and the definition of the MFAs and the corresponding shapefiles are available through the KCMD Dynamic Data Hub<sup>13</sup> for further analysis by scholars.

The authors would also like to acknowledge the GSMA,<sup>14</sup> colleagues from DG CONNECT<sup>15</sup> for their support and colleagues from Eurostat<sup>16</sup> and ECDC<sup>17</sup> for their input in drafting the data request.

<sup>13</sup> <https://bluehub.jrc.ec.europa.eu/catalogues/info/dataset/ds00170>.

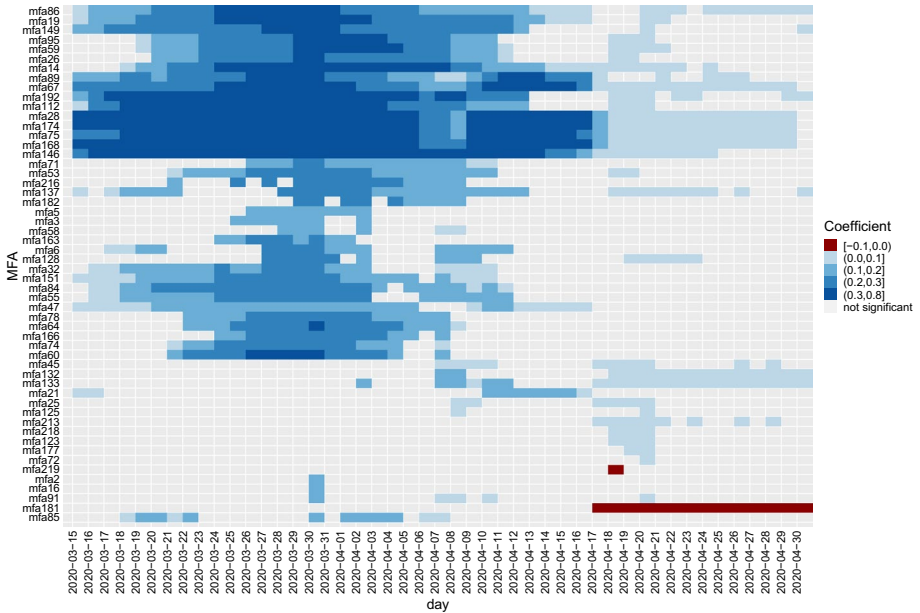
<sup>14</sup> GSMA is the GSM Association of Mobile Network Operators.

<sup>15</sup> DG Connect: The Directorate-General for Communications Networks, Content and Technology is the European Commission department responsible to develop a digital single market to generate smart, sustainable and inclusive growth in Europe.

<sup>16</sup> Eurostat is the Statistical Office of the European Union.

<sup>17</sup> ECDC: European Centre for Disease Prevention and Control. An agency of the European Union.





**Fig. 13** The impact of the MFA on the log number of cases. Blue dots mean that the corresponding MFAs is impacting (increasing) the number of cases; red dots means the contrary. transparent dots means that the corresponding regression coefficient is not significant. (Color figure online)

**Table 4** Elementary population size statistics (top) and number of COVID-19 cases (bottom) of the MFAs

Population	1st Qu.	Median	Mean	3rd Qu.
Positive impact MFAs (blue)	7943	31602	82666	101857
Not significant MFAs (transparent)	1183	3605	24494	15892
Negative impact MFAs (red)	6197	12211	12211	18225
Cases	1st Qu.	Median	Mean	3rd Qu.
Positive impact MFAs (blue)	441	1792	4508	5230
Not significant MFAs (transparent)	57	167	1118	713
Negative impact MFAs (red)	226	443	443	659

The labels correspond to the colors in Fig. 13

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

## Appendix

See Table 5.

**Table 5** Data used in the analysis of the MFAs. Some areas (like overseas territories) are excluded

Country	(ISO2)	highest granularity	used	NUTS 3	used	date range
Austria	(AT)	grid 16-18 km <sup>2</sup>	4789	districts	35	01/02/20–29/06/20
Belgium	(BE)	postal code areas	1131	regions	44	11/02/20–29/06/20
Bulgaria	(BG)	grid 16--18 km <sup>2</sup>	4615	provinces	28	01/02/20–27/06/20
Czechia	(CZ)	regular grid	4014	regions	14	01/01/20–28/06/20
Denmark	(DK)	municipalities	98	provinces	12	02/02/20–07/06/20
Estonia	(EE)	municipalities	79	counties	5	14/02/20–07/06/20
Spain	(ES)	municipalities	6893	provinces	59	01/02/20–29/06/20
Finland	(FI)	municipalities	310	provinces	19	02/02/20–07/06/20
France	(FR)	municipalities	1426	departments	96	01/01/20–23/06/20
Greece	(GR)	grid 25 km <sup>2</sup>	6240	prefectures	53	15/05/20 - 30/06/20
Croatia	(HR)	grid 17 km <sup>2</sup>	1384	counties	22	01/02/20 - 19/06/20
Italy	(IT)	census areas	8051	provinces	110	01/01/20 - 29/06/20
Norway	(NO)	municipalities	422	counties	18	02/02/20 - 07/06/20
Norway	(NO)	municipalities	356	counties	18	20/01/20 - 21/06/20
Sweden	(SE)	municipalities	290	counties	21	02/02/20 - 07/06/20
Slovenia	(SI)	grid 16--18 km <sup>2</sup>	1248	provinces	12	01/02/20 - 29/06/20

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

**Author contributions** SMI Manuscript writing, Content planning, Literature review, Modelling and Data Analysis. MV Content planning and Manuscript editing. SSL literature review, Manuscript editing, Maps building. CS, FS and DT Manuscript editing.

**Funding** Not applicable.

**Availability of material and data** Original mobile network data are not available and must be agreed individually with each MNO. The MFA's definitions and the corresponding shapefiles and mapping files are available at this link <https://bluehub.jrc.ec.europa.eu/catalogues/info/dataset/ds00170>.

**Code availability** All code written in R language (R Core Team 2020), available on request to the authors.

## Declarations

**Conflict of interest** No potential competing interest is reported by the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Andersen, A.K.: Are commuting areas relevant for the delimitation of administrative regions in Denmark? *Reg. Stud.* **36**(8), 833–844 (2002)
- Aurambout, J., Silva, B.E., Bosco, C.F., Conte, A., Ghio, D., Kalantaryan, S., Kompil, M., Perpiña, C.C., Proietti, P., Scipioni, M., Sulis, P., Tintori, G.: The demographic landscape of EU territories. Goujon, A., Jacobs, C., Natale, F., Lavalle, C. (eds.) vol JRC123046. Publications Office of the European Union, Luxembourg (2021). <https://doi.org/10.2760/49621>
- Ball, R.: The use and definition of travel-to-work areas in great Britain: Some problems. *Reg. Stud.* **14**(2), 125–139 (1980)
- Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. *J. Stat. Softw. Articles* **67**(1), 1–48 (2015). <https://doi.org/10.18637/jss.v067.i01>, <https://www.jstatsoft.org/v067/i01>
- Bates, D.M., DebRoy, S.: Linear mixed models and penalized least squares. *J. Multivar. Anal.* **91**(1), 1–17 (2004). <https://doi.org/10.1016/j.jmva.2004.04.013>, <https://www.sciencedirect.com/science/article/pii/S0047259X04000867>, special Issue on Semiparametric and Nonparametric Mixed Models
- Bwambale, A., Choudhury, C., Hess, S., Iqbal, M.S.: Getting the best of both worlds: a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling. *Transportation* (2020)
- Casado-Díaz, J.M.: Local labour market areas in Spain: a case study. *Reg. Stud.* **34**(9), 843–856 (2000)
- Csáji, B.C., Browet, A., Traag, V.A., Delvenne, J.C., Huens, E., Van Dooren, P., Smoreda, Z., Blondel, V.D.: Exploring the mobility of mobile phone users. *Physica A* **392**(6), 1459–1473 (2013)
- Dijkstra, L., Poelman, H., Veneri, P.: The EU-OECD definition of a functional urban area (2019). Available at <https://doi.org/10.1787/d58cb34d-en>
- EDPB.: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the covid-19 outbreak (2020). Available at <https://bit.ly/2DegEJz>
- European Commission.: Commission recommendation (EU) on a common union toolbox for the use of technology and data to combat and exit from the covid-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data, 2020/518 (2020a). Available at <http://data.europa.eu/eli/reco/2020/518/oj>
- European Commission.: The joint European roadmap towards lifting covid-19 containment measures. (2020b). Available at <https://bit.ly/2DiOj4Y>
- Eurostat.: Territorial typologies manual - cities, commuting zones and functional urban areas. Eurostat (2016). Available at <https://bit.ly/3itCj0i>
- Fekih, M., Bellemans, T., Smoreda, Z., Bonnel, P., Furno, A., Galland, S.: A data-driven approach for origin-destination matrix construction from cellular network signalling data: a case study of Lyon region (France). *Transportation* (2020). <https://doi.org/10.1007/s11116-020-10108-w>
- Fisher, R.: Statistical methods for research workers. *J. R. Stat. Soc.* **89**(1), 144–145 (1926). <https://doi.org/10.1111/j.2397-2335.1926.tb01837.x>
- Fred, A.L., Jain, A.K.: Combining multiple clusterings using evidence accumulation. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(6), 835–850 (2005)
- Gabrielli, L., Fadda, D., Rossetti, G., Nanni, M., Piccinini, L., Pedreschi, D., Giannotti, F., Lattarulo, P.: Discovering mobility functional areas: a mobility data analysis approach. In: Cornelius, S., Coronges, K., Gonçalves, B., Sinatra, R., Vespignani, A. (eds.) *Complex Networks IX*, *CompleNet* 2018, pp. 311–322. Springer, Cham (2018)
- Gelman, A., Hill, J.: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press (2006)
- Gravilov, M., Anguelov, D., Indyk, P., Motwani, R.: Mining the stock market: Which measure is the best? Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining, pp. 487–496 (2000)
- GSMA.: The mobile economy 2020 report (2020). Available at <https://www.gsma.com/mobileeconomy/>
- Iacus, S.M., Santamaria, C., Sermi, F., Spyrtos, S., Tarchi, D., Vespe, M.: Human mobility and covid-19 initial dynamics. *Nonlinear Dyn.* (2020). <https://doi.org/10.1007/s11071-020-05854-6>

- Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N.A.: Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* tba:1–11 (2020)
- Killer, V., Axhausen, W.: Mapping overlapping commuting-to-work areas. *J. Maps* **6**(1), 147–159 (2010)
- Kraemer, M.U., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M., du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., et al.: The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**(6490), 493–497 (2020)
- Van der Laan, L.: Changing urban systems: an empirical analysis at two spatial levels. *Reg. Stud.* **32**(3), 235–247 (1998)
- Mamei, M., Bicocchi, N., Lippi, M., Mariani, S., Zambonelli, F.: Evaluating origin-destination matrices obtained from cdr data. *Sensors* **19**, 1440 (2019)
- Novak, J., Ahas, R., Aasa, A., Silm, S.: Application of mobile phone location data in mapping of commuting patterns and functional regionalization: a pilot study of estonia. *J. Maps* **9**(1), 10–15 (2013)
- OECD.: Redefining Territories. OECD (2002). Available at <https://doi.org/10.1787/9789264196179-en>
- Patterson, Z., Farber, S.: Potential path areas and activity spaces in application: a review. *Transp. Rev.* **35**(6), 679–700 (2015). <https://doi.org/10.1080/01441647.2015.1042944>
- Pons, P., Latapy, M.: Computing communities in large networks using random walks. *J. Graph Algorithms Appl.* **10**(2), 191–218 (2006)
- R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2020). <https://www.R-project.org/>
- Wesolowski, A., Eagle, N., Tatem, A.J., Smith, D.L., Noor, A.M., Snow, R.W., Buckee, C.O.: Quantifying the impact of human mobility on malaria. *Science* **338**(6104), 267–270 (2012)
- Wu, J., Leung, K., Leung, G.: Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *Lancet* **395**(10225), 689–697 (2020)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Stefano Maria Iacus** is a project officer at the European Commission, Joint Research Centre, previously full professor of Statistics at the University of Milan (1999–2021). Holds a PhD in Statistics from University of Padua (IT) and Doctorate en Mathématiques from University of Le Mans (FR). Former R-core member (1999–2017) and now R Foundation Member. His research interest include simulation and inference for stochastic processes, causal inference, sentiment analysis, computational statistics, forecasting human migration and demographic modelling. At the JRC he leads the team on data innovation for migration and demography.

**Carlos Santamaria** holds an electrical engineering degree from the Technical University of Catalonia, Spain (2001), and an MBA from the Technical University of Madrid, Spain (2008). His research has focused on remote sensing, computer vision, machine learning, maritime surveillance, and human migration and mobility, both for policy support and for industrial applications.

**Francesco Sermi** is a scientific researcher and data analyst at the Joint Research Centre (JRC). He is part of the Knowledge Centre on Migration and Demography (KCMD). He applies his background in telecommunication engineering, signal and data processing, and radar design to data analysis and visualisation in the field of migration and demography. He is behind both the KCMD Atlases: the Atlas of Migration (in Dec 2021 the 4th release) and the Atlas of Demography (launched in April 2021). He also works on the New Pact on Migration and Asylum and on Big Data and Human Mobility. His researches have been published in several international journals.

**Spyridon Spyrtatos** is a researcher at the Knowledge Center on Migration and Demography of the European Commission. His research focuses on human geography, demography and migration using innovative data and data science methods. He also worked at the Joint Research Center of the European Commission as a trainee (2011–2012) and as a PhD researcher (2012–2015). He holds an MSc degree in Geographical Information Science from University College London, UK (2011), a BSc degree in Urban and Regional Planning (2010) and a PhD in the use of crowd-sourced data for Urban analysis (2017) from the University of Thessaly, Greece.

**Dario Tarchi** is the Deputy Head of Unit, Demography, Migration and Governance Unit, Joint Research Centre, European Commission. Dario Tarchi is a physicist by training. He has been with the JRC for more than 20 years. He has been involved, as a Project Leader, in the design, creation and operation of the Knowledge Centre on Migration and Demography since its establishment in 2016. His main responsibilities include designing and implementing knowledge management tools for evidence-informed policymaking, notably the Atlas of Migration and, more recently, the Atlas of Demography.

**Michele Vespe** is a scientific officer at the European Commission, Joint Research Centre (JRC), where he coordinates a team of researchers investigating societal consequences associated with the improved availability of digital trace data. This includes research in the fields of data governance and computational social science. He was previously a Senior Scientist with NATO and a project engineer in industry. He holds a telecommunications engineering degree from University of Florence (2003) and a PhD in signal processing from University College London (2006). He is part of the editorial board of the Data & Policy journal.