Mobility and COVID-19 in Andorra: Country-Scale Analysis of High-Resolution Mobility Patterns and Infection Spread

Ronan Doorley, Alex Berke, Ariel Noyman, Luis Alonso, Josep Ribó, Vanesa Arroyo, Marc Pons, and Kent Larson

***Abstract*—Throughout the COVID-19 pandemic, nonpharmaceutical interventions, such as mobility restrictions, have been globally adopted as critically important strategies to curb the spread of infection. However, such interventions come with immense social and economic costs and the relative effectiveness of different mobility restrictions are not well understood. Some recent works have used telecoms data sources that cover fractions of a population to understand behavioral changes and how these changes have impacted case growth. This study analyzed uniquely comprehensive datasets in order to examine the relationship between mobility and transmission of COVID-19 in the country of Andorra. The data consisted of spatio-temporal telecoms data for all mobile subscribers in the country, serology screening results for 91% of the population, and COVID-19 case reports. A comprehensive set of mobility metrics was developed using the telecoms data to indicate entrances to the country, contact with tourists, stay-at-home rates, trip-making and levels of crowding. Mobility metrics were compared to infection rates across communities and transmission rate over time. All metrics dropped sharply at the start of the country's lockdown and gradually rose again as the restrictions were gradually lifted. Several of these metrics were highly correlated with lagged transmission rate. There was a stronger correlation for measures of indoor crowding and inter-community trip-making, and a weaker correlation for total trips (including intra-community trips) and stay-at-homes rates. These findings provide support for policies which aim to discourage gathering indoors while lifting the most restrictive mobility limitations.**

***Index Terms*—Cellular phones, geospatial analysis, health and safety, Internet of Things, mobile communication.**

Manuscript received March 26, 2021; revised September 7, 2021; accepted October 10, 2021. Date of publication October 19, 2021; date of current version January 5, 2022. This work was supported by the Andorra Innovation Hub provided financial support to the City Science Group of the MIT Media Lab. *(Corresponding author: Ronan Doorley.)*

Ronan Doorley, Alex Berke, Ariel Noyman, Luis Alonso, and Kent Larson are with the City Science Group of the Media Lab, MIT, Cambridge, Massachusetts, MA 02139-4307 USA (e-mail: doorleyr@mit.edu; aberke@mit.edu; noyman@media.mit.edu; alonsolp@media.mit.edu; kll@mit.edu).

Josep Ribó, Vanesa Arroyo, and Marc Pons are with the Andorra Innovation Hub, AD500 Les Escaldes, Andorra (e-mail: jribof@actuatech.ad; vanesa.arroyo@actuatech.ad; marc.pons@actuatech.ad).

Digital Object Identifier 10.1109/JBHI.2021.3121165

I. Introduction

THE rapid spread of the COVID-19 pandemic in 2020 led to an unprecedented global effort to curb the disease. Given the lack of effective therapeutics or widely distributed vaccines at the start of the pandemic, Nonpharmaceutical Interventions (NPIs) became a primary public health strategy for achieving these goals [10], [24]. These included widespread government mandates restricting mobility and economic activity, in ways that were unimagined in months previous to 2020.

Even after some populations have been largely vaccinated against COVID-19, NPIs may continue to be considered as important strategies to manage new variants of COVID-19 and future health crises.

Economic lockdowns, border closures and mobility restrictions are clearly effective to some degree in curbing transmission of COVID-19 [6]. However, these measures have enormous economic, social and mental health impacts and therefore policy-makers must carefully consider the relative costs and benefits of every policy. Unfortunately, neither the costs nor benefits of NPIs for COVID-19 are well understood due the enormous complexity of the problem in terms of both virology and human dynamics.

Until recently, gathering fine grained movement patterns data for large populations of people was nearly impossible. With the advent of GPS-enabled smartphones, most people around the world today are continuously generating data on their mobility patterns. In particular, telecoms companies routinely record histories of device connections to cell towers for billing purposes, and these data can be used to reconstruct individual device trajectories.

Numerous studies that use mobile phone data to analyze human mobility and behaviour have been published over the past several years [18], [20], [33]. In the past year, researchers have also used mobile phone data to study mobility in the context of the COVID-19 pandemic. Some studies only look at how behaviour has changed and do not incorporate data on COVID-19 infections [9], [25], [34]. Other studies use mobility data as inputs to build metapopulation compartmental models, which divide the population into smaller communities, assuming reduced transmission between communities [5], [26], [43].

Many of these works assume that disease transmission is independent of the density of people, whereas other studies assume that human interactions, and therefore disease transmissibility, are proportional to density of people [11]. A good understanding

of the influence of density of people on transmission rates is essential as the chosen assumption can be expected to influence the results of models and simulations.

Related works that use mobility data in COVID-19 analyses have relied on data collected from only a fraction of the population, or used only precomputed mobility metrics, such as the number of trips to points of interest (POIs). Such data sources include Facebook's Data for Good program [19], Google's Community Mobility Reports [21], and other mobility data aggregators [39], [42].

For example, researchers estimated the reproduction rate for COVID-19 in multiple countries and partially validated their results by checking the correlation between the time series of their estimates and a mobility index computed from Google's Community Mobility Reports data [8]. (Note that Google's mobility data is not available for Andorra.)

In contrast to these related works, this study makes use of telecoms data that covers *all* mobile phone users in the study area in order to develop a comprehensive set of mobility metrics. These metrics include country entrances, stay-at-home rates, trips, and crowding. This data is used to analyze the population's mobility behaviors before and after the government lockdowns.

In addition to telecoms data, this study makes use of COVID-19 antibody test results covering over 91% of the population. The combination of full coverage telecoms data and antibodies data provided a unique opportunity to study the spread of COVID-19 in a relatively closed population.

The goal of this study is to empirically examine the evolution of mobility behavior and densities of people over time and space and their relationship to the transmission rate of COVID-19. This analysis aims to provide decision support to policy-makers considering the use of NPIs. In addition, empirical observations of how inter-community travel and population density have affected COVID-19 transmission rates can inform the development of models for future infectious disease crises.

This study is based in the country of Andorra and explores the relationships between government policies, several inferred mobility metrics and the nature by which COVID-19 spread across the country. More specifically, this work attempts to answer the following questions: How did travel patterns change before, during and after lockdown policies were implemented? Did changes in mobility behaviour and crowding over time correspond to changes in transmission rate over time? Which changes in mobility behavior are most closely correlated with changes in transmission rate?

Section II provides context for the work by briefly introducing the COVID-19 pandemic in Andorra. The data sources, pre-processing and mobility metrics used in the study are then described in Section III. Section IV presents analysis and findings about the relationship between mobility and COVID-19 infections in Andorra. The implications of the findings are discussed in Section V, followed by a conclusion.

II. Andorra and COVID-19

The study region for this work was the European microstate of Andorra. Andorra is located in the heart of the Pyrenees mountains, situated between the borders of France and Spain. It lacks an airport or train service so the primary way to enter or exit the country is by crossing the French or Spanish border by car (see Fig. 1). For this reason, the stringent border closures by France and Spain during the pandemic effectively closed the borders of Andorra.

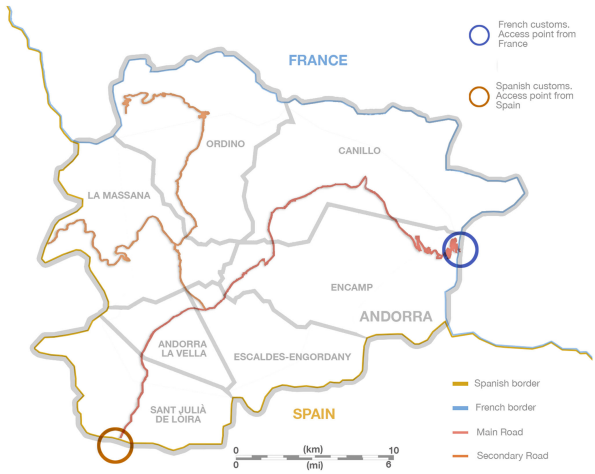

Fig. 1. The country of Andorra is situated between France (north) and Spain (south) and divided into 7 parishes. The main entry routes to the country, and roads between different parishes, are highlighted.

Andorra has a population of approximately 78,000 people and receives about 8 million visitors in a normal year [15]. The country is divided into 7 parishes which are shown in Figure 1. On March 2, 2020, the country's first coronavirus case was confirmed. Announcements of lockdown measures beginning on March 13, as well as border closures [12], [32], soon followed. A list of key dates and policies is given in Table I.

III. Data and Methodology

Three primary data sources were used in this study: (i) the results of COVID-19 serology testing in which more than 91% of the population participated, (ii) daily reported COVID-19 case counts for Andorra and (iii) the telecoms data for all mobile subscribers in Andorra. The telecoms data were used to develop and analyze metrics that are indicative of people's mobility and behaviors. The metrics include the daily number of entrances to the country, the fraction of the population staying home, the number of people making trips between the country's regions (parishes), and measures of crowding. In addition, serology and reported cases data were used to analyze infection rates and the temporal relationship between infection rates and mobility behaviors. Each of the data sources and the computation of metrics are described in the sub-sections below. The code for the analysis is publicly available in a Github repository: https://github.com/CityScope/CSL_Andorra_COVID_Public.

A. Reported Cases Data

This analysis uses daily confirmed COVID-19 cases reported from March 1st to October 31st of 2020. This period spans the time from the first reported case in Andorra to the end of the period of mobility metrics used in this study. The data source is the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [17].

TABLE I
TIMELINE OF THE COVID-19 RELATED DATES IN ANDORRA

March 2	First coronavirus case confirmed in Andorra.
March 13	Partial lockdown by Andorran government.
March 17	French border closure.
March 18	Total lockdown by Andorran government.
March 19	Spanish border closure.
April 7	Beginning of masks delivery and progressive use by population.
April 17	Allowance of 1 hour of walk in 1km radius every two days.
April 20	Phase 1 reopening. Low risk economic activities resume. 1000 people return to work.
May 1	Start of 1st round of population serology screening.
May 4	Phase 2 reopening. Additional 4760 workers return to normal activity.
May 13	Increase to 2 hours of activity every day to walk or exercise.
May 14	Start of 2nd round of population serology screening.
May 18	Phase 3 reopening. Additional 3300 workers return to normal activity.
June 1	Confinement restrictions completely lifted.
June 15	French and Spanish borders are open (with some restrictions on the Spanish side).
June 21	Neighboring country, Spain, lifts state of emergency and lifts some border controls.
July 1	Remaining Andorran restrictions on Spanish border are lifted.
Sept. 1	Massive testing for teachers and school children.
Oct. 7	Bars and restaurants access is restricted.
Oct. 10 - 12	Massive entry of Spanish tourists due to Spanish holiday.
Oct. 22	Massive testing strategy to test 40% of the population every week begins.
Oct. 26	Restrictions on movement between both the Spanish and French borders.
Oct. 28	Restrictions on in-person meetings. Working at home is mandatory if possible.

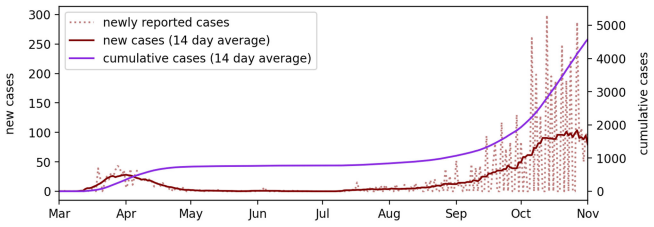

Fig. 2. A timeline of new and cumulative case reports. The dotted line shows the daily new case reports. The solid lines show values smoothed over a 14-day rolling average.

The reported cases data are subject to error due to reporting lags and irregular testing periods. In particular, there was an influx of late case reports in early June which were identified through the mass serology testing [22]. These reports were removed for the purposes of temporal analysis because they were not registered at the actual date of testing, and furthermore they resulted from testing which was not carried out at other time periods and would therefore bias the temporal analysis. To mitigate reporting irregularities, this analysis smooths the case reports data by averaging over a rolling window of 14 days (see Fig. 2).

B. Serology Data

In May, 2020, the Andorran government launched a campaign to provide voluntary serology testing for its entire population. The screening was conducted for a previous research study [38], which was approved by the relevant Andorran agencies and local research ethics committee. As shown in Fig. 2, during this period and the preceding weeks, the cumulative case count was relatively flat, implying that the number of people with either current or prior infections remained stable during this period. Participants were tested for presence of both IgM and IgG antibodies twice, with two weeks between each round of testing. 91% of the population completed the first round of testing and 86% completed the second round. The test data were annonymised, so that individual-level analysis of risk was not possible. However, the data included additional variables, including parish of residence and residency status, allowing group infection rates to be compared. Livzon IgM/IgG Diagnostic Kits were used [30]. All such screenings are subject to some level of error, with false positives and false negatives. Independent testing of the kits had previously been carried out to establish the sensitivity and specificity of the tests. Maximum likelihood estimation was used to estimate the actual number of people with current or previous infections, accounting for sensitivity and specificity [37]. Having estimated the overall prevalence, the probability of every individual having a current or previous infection was calculated using Bayes' Theorem [31].

C. Telecoms Data

For this study, telecoms data were provided by Andorra Telecom (AT). Since AT is the sole telecoms provider in Andorra, the data cover all mobile subscribers in the country, unlike most telecoms datasets where the market is fragmented. Each observation in the AT data includes a unique ID for the subscriber, a timestamp, the coordinates of the device at the time of the update, and nationality for the subscriber's home network. The coordinates are given in the standard WGS84 system and estimated by AT to be accurate to within 25-100m in urban areas. AT utilizes enhanced triangulation using the angle, time and RSS received by the device from different base-stations as well as error mitigation using a proprietary algorithm. The AT data have been further described in [18], [33]. In order to comply with data protection policies, the telecoms data were anonymised and only used to compute metrics which contained no identifying information. Any attributes associated with metrics, such as parish of residence, represent groups of several thousand people at least.

This study used data from March 2019 to June 2019 and January 2020 to October 2020, from both the 3G and 4G networks. Data recorded by AT were missing or incomplete in periods of the study (Feb 14 - Mar 1, Jun 28 - Jun 29, Jul 21 - Jul 27, and Oct 1 - Oct 6, 2020). These gaps resulted in missing values in the computed mobility metrics, and are displayed as gaps in the time series plots of these metrics.

1) Pre-Processing: During study period, a total of 1,288,799 unique subscriber IDs were observed in the AT data, most of whom were tourists. As described below, the data for all subscribers were pre-processed in order to identify stay-points, home parishes, days present in Andorra and residency status.

To reduce noise, and increase the probability of actual location, the series of location observations for each subscriber was reduced to a series of stay-points of 10 minutes or more within a radius of 200 m or less. The stay-point extraction algorithm of Li *et al.* (2008) [29] was used for this purpose. In order to

compute daily mobility metrics across all people present in the country, it was necessary to know which people were present on any day. Some devices were found to be unobserved in the data for several days, even during the full government lockdown. This may be due to a combination of inactivity, lack of telecoms reception in certain areas, and/or noisy data. It was assumed that gaps in data of two weeks or more represented true absence from the country. The beginnings and endings of periods of presence were counted as entrances to and departures from the country respectively.

By observing the distribution of total number of days present across all subscribers, it was found that the majority of subscribers were present fewer than 50 days. The rest of the subscribers were present for a significantly greater number of days. 50 days of presence in Andorra was therefore considered as an appropriate cutoff for identifying tourists. Further analysis uses this categorization of tourists versus non-tourists.

Home parishes: Home parishes were inferred for each subscriber on a monthly basis, since some residents of Andorra moved during the periods of study and since the tourist population fluctuated. To infer home parish, each stay-point first was assigned to the parish in which it was contained. Each subscriber's home parish was then determined to be the parish in which they spent the most cumulative time during night-time hours (12:00 AM to 6:00 AM). This method is similar to those employed in related studies of human mobility that use cellular data [34], [35]. To evaluate the representativeness of the telecoms data and this methodology, the distribution of home parishes inferred for subscribers in the last month of the lockdown (May, 2020) was compared to the projected 2020 population statistics published by the Andorran government [15]. The Pearson correlation coefficient was found to be 0.959, suggesting that the telecoms data are highly representative of the true population.

2) Mobility Metrics: A number of key mobility metrics were computed based on the AT data and used for this report. The metrics are described below.

Stay-at-Home Rates: Advising people to stay at home as much as possible is widely recognised as an effective policy for reducing transmission of infection. The stay-at-home rate for a parish on a given day was defined as the number of residents of that parish present and staying home divided by the total number of people present. A person was assumed to have stayed at home if they had at most one stay-point in their own parish and no stay points in other parishes. This assumes that anyone making a trip would have recorded at least two stay-points as they moved between cell coverage areas.

Trips: One purpose of lockdown policies was to reduce the population's mobility, and the risk of a mobile population potentially carrying disease between places they go. The risks associated with trip making may be particularly high when trips occur between two communities at different stages of the epidemic curve as this can introduce contact between people of relatively high infection probability and a population with high susceptibility. For this reason, trips between parishes are of particular interest. Metrics used in this analysis included (i) the number of total daily trips, (ii) total daily trips between parishes, (iii) daily number of distinct subscribers making any trips and (iv) daily number of distinct subscribers making trips between parishes.

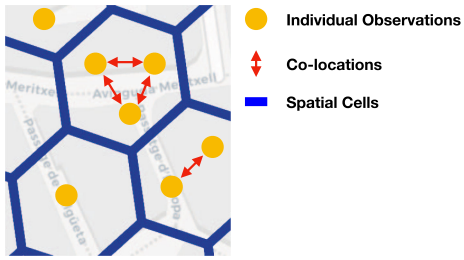

Fig. 3. The computation of crowding index based on individual spatial observations contained in H3 cells. The crowding index for one spatio-temporal cell is the count of unique person-person pairs co-located in the cell.

The number of daily trips for each subscriber was computed as their daily number of stay points minus 1, since a new stay point is recorded when a subscriber moves beyond a 200 m radius. Trips between parishes were counted when consecutive stay points occurred in different parishes. The total trips metrics were computed as a sum over each users' trips.

Entrances: Entrances of tourists or other travellers to the country creates the risk of importing cases from other countries with higher incidence, as shown in a recent study [23]. The calculation of number of new entrances to the country on each day is described in Section III-C1.

Crowding Indices: Dense crowds of people create high potential for human interaction and transmission, especially indoors. As a proxy for such crowding, a set of crowding indices were defined as follows. The time period and study region were first divided into discrete spatio-temporal intervals. H3 cells of resolution 11 [1] were used as the geographic intervals. At this resolution, the H3 cells have an approximate diameter of 50 m [4] which roughly corresponds to the resolution of the telecoms data in urban areas. Time was divided in 20 minute intervals. In every spatio-temporal interval, the number of unique person pairs co-located in the interval was counted to find the crowding index for that spatio-temporal cell. This is illustrated in Fig. 3. This value was aggregated over all cells to calculate the crowding index for the entire study area. Additional crowding indices representing the degree to which certain categories of people crowed together were also calculated by only counting specific person pairs, as described in equations 1. Additionally, a distinction was made between indoor crowding and outdoor crowding based on the percentage of indoor and outdoor area in the H3 cell. These percentages were calculated using the Global Human Settlement Layers (GHSL) [13]. The crowding indices can be defined as below:

$$
\begin{aligned}
CI_{tj} &= \frac{N_{tj} \times (N_{tj} - 1)}{2} \\
CCI_{tj}^{x,y} &= \frac{N_{tj}^{x} \times N_{tj}^{y}}{2} \\
ICI_{tj} &= \delta_j \times CI_{tj} \\
DCCI_{d}^{x,y} &= \sum_{j \in J, t \in D} CCI_{tj}^{x,y} \\
DICI_{d} &= \sum_{j \in J, t \in D} ICI_{tj}
\end{aligned} \tag{1}
$$

Where
CI_{tj} = Crowding Index: the number of unique person pairs co-located in cell j during time interval t.
N_{tj} = the number of subscribers observed in cell j during time interval t.
$CCI_{tj}^{x,y}$= Cross Crowding Index: the number of unique pairs of a person of category x co-located with a person of category y in cell j during time interval t.
N_{tj}^{c} = the number of subscribers of category c observed in cell j during time interval t.
ICI_{tj} = Indoor Crowding Index: the number of unique person pairs co-located in cell j during time interval t, attributable to indoor areas.
δ_j = the fraction of land area of cell j covered by buildings.
$DCCI_{d}^{x,y}$ = Daily Cross Crowding Index: the aggregation of $CCI_{tj}^{x,y}$ d over all spatio-temporal cells during day d.
DCI_d = Daily Crowding Index: the aggregation of CI_{tj} over all spatio-temporal cells during day d.
$DICI_d$ = Daily Indoor Crowding Index: the aggregation of ICI_{tj} over all spatio-temporal cells during day d.
D = the set of all intervals t in day d.
J = the set of all spatial cells in the study area.

The purpose of the crowding indices was not to precisely count the numbers of interactions between people as the resolution of the data would not allow for this. The aim was rather to compute indices which are generally indicative of crowding, which could be compared to measures of the transmission rate of COVID-19. In order to test the robustness of the crowding indices to the choice of cell resolution, DCI_d was computed for every day in the study period using cell resolutions of 10, 11 and 12. The resulting time series for cell resolution of 11 had correlations of 0.97 and 0.98 with the series for resolutions of 10 and 12 respectively. These high correlations indicate that the choice of cell resolution only impacts the scale of the crowding index.

A similar metric was also computed at the individual level in order to study the distributions of co-location rates among individuals. Each day, a person-to-person co-location graph, G_d was defined as follows. A node was created for every person present in the country that day. A pair of nodes was linked if those two people had been co-located in one or more spatio-temporal intervals during that day. The degree of each node was defined as the number of Daily Co-locations of the corresponding person.

IV. Analysis and Findings

A. Infections

In the first round of serology testing, IgM and IgG antibodies were detected in 9.6% of participants. In the second round of testing, this fell to 8%. In addition, participation rates fell from 91% to 86% between rounds. The reduction in antibody detection suggests that the antibodies fell below the detection limit rapidly. For this reason, as well as the reduction in test participation, the first round results alone were considered as a better indicator of overall exposure, while still being conservative. Using maximum likelihood estimation as described in Section III, the overall infection rate was estimated to be 11%. By comparing the number of reported cases before May 1st to the inferred number of infections, it was estimated that just 1 in 11 cases of COVID-19 were recorded in official statistics before this date.

TABLE II
Populations, Participant Numbers And Aggregated Infection Rates by Parish Estimated from the May 2020 Serology Testing Program in Andorra. Note that the Number of Participants in Canillo Was Greater than the Official Population Due to the High Concentration of Temporary Workers

Parish	Population	Participants	Infection Rate
Andorra la Vella & Escaldes-Engordany	37066	33361	0.1
Canillo	4371	4489	0.11
Encamp	14626	10830	0.1
La Massana	10199	8829	0.14
Ordino	4957	4205	0.11
Sant Juliá de Lória	9374	8612	0.13

The probabilities of each individual having a current or prior infection during the first round were then calculated using Bayes' Rule [31]. These probabilities were then averaged across residence parishes and residency status to estimate the percentage of infections in each group. Note that the adjacent parishes of Andorra La Vella and Escaldes-Engordany were combined for the serology analysis. This was done on the advice of Andorran officials because the two parishes comprise a single urban area and were sometimes interchanged in the address field of the serology data. The populations and infection rates in each parish are shown in Table II. The overall infection rate for temporary workers was estimated to be 14% compared to 11% for ordinary residents. La Massanna was the parish with the highest overall infection rate of 14% compared to the lowest rate of 9% in Andorra la Vella. In order to determine whether the differences in infection rate across residency status and parish of residence were significant, chi squared tests were performed. It was found that both residency status and parish of residence were significantly associated with infection rate ($p < 1e\text{-}5$).

B. Mobility

The mobility of the population is reported on in several phases. Firstly, in order to understand the initial outbreak, mobility patterns during January and February 2020 were analysed with an emphasis on tourists. Mobility patterns from March onward were then analyzed in order to assess how the population responded to the pandemic and to government policies. The distribution of mobility behaviours is also analyzed in order to explore whether highly active people and events may have disproportionately contributed to transmission risk. The evolution of mobility behaviour over time is then compared to the transmission rate during this period.

1) Tourism before Outbreak: The pre-outbreak analysis was focused on the period from January 1st to February 14th, 2020. Unfortunately, the telecoms data for the last 2 weeks of February were unavailable. The first recorded case of COVID-19 in Andorra was on March 2. January and February are part of the ski season in Andorra and ski resorts were identified as high risk environments for the spread of COVID-19 in Europe [2], [3], [14]. Mobility of tourists and contact with tourists were therefore key considerations. The largest number of tourists during this period (45%) were found to have stayed in the parish of Canillo.

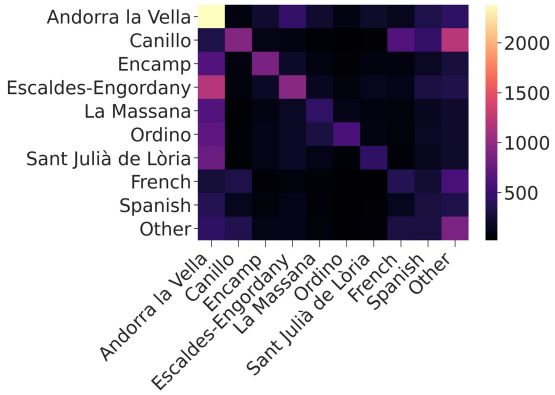

Fig. 4. Average Daily Cross Crowding Index (DCCI) between residents of each parish and tourists by country of origin. The value of row i, column j corresponds to the DCCI between groups i and j, normalised by the population of group i. Residents of Canillo showed a high degree of crowding with tourists.

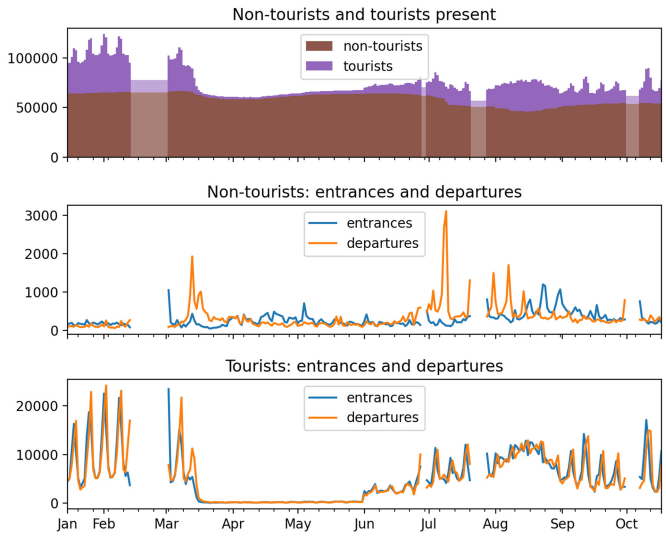

Fig. 5. Andorra telecoms data were used to estimate the daily number of country entrances, departures, and population present, for tourists and non-tourists. Metrics are shown for Jan.–Oct. 2020. Gaps in the available data during the study (Feb. 14–Mar. 1, Jun. 28–Jun. 29, Jul. 21–Jul. 27, and Oct. 1–Oct. 6, 2020) resulted in gaps in the time series plots of the metrics and in spikes in inferred departures and entrances directly before and after the gaps, respectively.

In order to more directly assess the levels of contact between tourists and residents of each parish, mobile subscribers were categorised as either a resident of one of the seven parishes or as a tourist. Tourists were further categorised as Spanish, French or Other based on their home network. The average Daily Cross Crowding Index between each of these 10 groups was computed as described in Section III-C2. As shown in Fig. 4, observed co-locations with tourists were much higher for residents of Canillo than for any other parish.

2) Entrances and Departures: Throughout the pre-lockdown and lockdown periods, entrances into and departures from the country corresponded with border restrictions. As shown in Fig. 5, this behavior differed by residency status. As expected, the initial lockdown policies that began March 13th,

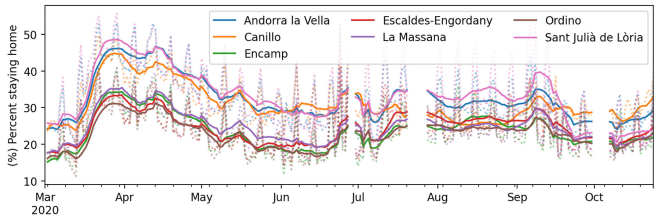

Fig. 6. Telecoms data were used to estimate stay-at-home rates for each of the 7 parishes in Andorra. The dotted lines show the estimated daily percent of residents staying home and solid lines show 7-day rolling averages.

and the border closures by neighboring countries, led to mass departures from Andorra. This is observed in the telecoms data as a large drop-off in the number of subscribers present in mid-March. Many of these subscribers were tourists, and some may also have been temporary workers. The population was relatively stable between mid March and the end of May while lockdown was in effect and serology testing was carried out. In early May, there was a small surge of non-tourists arriving to the country and towards the end of May, there was a small surge of non-tourists leaving the country. This is shown in Fig. 5. These observations may reflect people arriving in order to participate in serology testing as well as leaving the country after receiving one or both test results. Further analysis found that the surge in departures by non-tourists at the end of May was driven by people who did not return to the country.

On June 1st, the lockdown in Andorra was lifted and both tourists and non-tourists began to enter the country again. Initially, the tourists were mostly French, followed by an influx of Spanish tourists later in June, corresponding to the reopening of Spanish borders on June 21st. There was a spike in departures by non-tourists in the second weekend of July (July 9 saw over 3100 departures), corresponding to the weekend at the start of the common holiday period of Andorrans.

3) Lockdown Mobility Behaviour by Parish: The stay-at-home rates were compared for each parish over time in order to determine whether adherence to social distancing policies differed by parish and whether such differences corresponded to differences in infection rates by parish.

As shown in Fig. 6, all parishes exhibit a similar trend with a sharp increase in the portion of people staying home at the start of the most restrictive lockdown measures in mid March. There is then a gradual drop off as lockdown policies were relaxed. At the start of the lockdown period, there were two distinct clusters of parishes in terms of the average stay-at-home rate: Andorra la Vella, Canillo, and Sant Juliá de Lória exhibited the highest portion of users staying home. However, this difference gradually diminished as the months continued. These differences in stay-at-home rates by parish have no clear correlation with infection rates by parish.

4) Distributions of Co-locations: The distributions of co-locations were analyzed in order to determine the extent to which potentially risky mobility behaviours were dominated by relatively small numbers of people and events. This would be the case if the distributions of the mobility metrics had high positive skew or “fat-tails”.

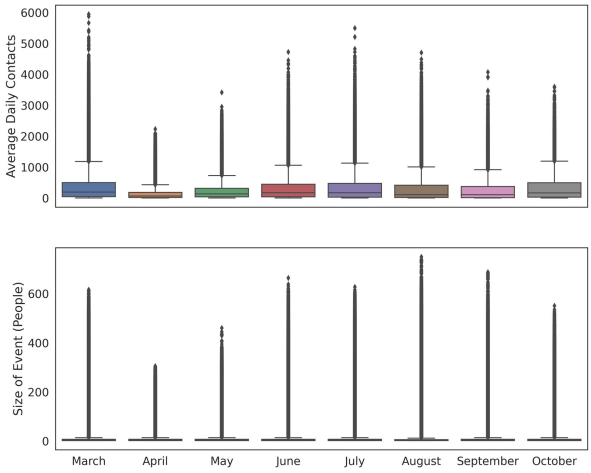

Fig. 7. Box plots showing distributions of (Top) average daily co-locations per person for 3-day periods and (Bottom) "event" sizes, in each month. Both measures show high positive skew in each month indicating small numbers of people and events accounted for disproportionately high levels of co-locations.

In order to evaluate how evenly or unevenly co-locations were distributed across people, and how this changed over time, the daily co-location graphs described in Section III-C2 were used. The distributions of Daily Co-locations were evaluated for a sample of 3 days in each month from March 2020 to October 2020 (the first available Monday-Wednesday period) in order to smooth out unusual behaviours on any one day. As shown in Fig. 7, the distribution of number of co-locations per individual had a high positive skew in every period considered. Co-locations were highest in March (before the lockdown) and lowest in April. For each month, the individuals were then ranked by their Daily Co-locations and the top 20% of users were removed from the daily graphs. It was found that by removing these top 20% most "socially active" active people in each month, between 78% (May) and 85% (August) of the connections were eliminated.

In order to evaluate the contribution of potential super-spreader events, an event was defined as any spatio-temporal interval between March and October 2020 where at least 2 people were co-located, as defined in Section III-C2. The distribution of event sizes was evaluated for each month. As shown in Fig. 7, the number of people-per event showed high positive skew in each month. The distribution of event sizes contracted significantly in April and to a lesser extend in May and October. The crowding index, CI_{tj}, for every event was computed as described in Section III-C2 and it was found that the top 20% of events across the whole period were responsible for 96% of all co-locations.

C. Mobility and Transmission Rate

In this section, population behaviors, estimated as mobility metrics, are compared to disease transmissibility metrics.

1) Transmission Rate: The transmission rate of an infectious disease, β can be simply defined as the average number of people an infected person would infect per unit time, if everyone they met were susceptible [7]. This metric is time-varying and is impacted by a population's behaviors. In standard SIR modelling, the rate of occurrence of new cases is equal to the product of the transmission rate, the number of currently infectious people and the fraction of the population who are still susceptible [26].

Therefore, a simple proxy for the transmission rate per day can be estimated by normalising the new infections per day by the number of currently infectious people and by the fraction of the population who are susceptible. A common approach to normalise new infections by the current number of infectious people is to take the log case growth over a suitable period, such as 2 weeks [36], [40]. The susceptible fraction is more difficult to estimate without good data on reporting rates. However, this study's access to a combination of serology testing and case reports made it possible to estimate a reporting factor. Furthermore, due to the availability of high resolution telecoms data, a real-time population census was also available. This allowed a running estimate of the susceptible fraction to be computed. Note that the assumption of a constant reporting rate may have led to an overestimation of the susceptible population in the latter part of the study period. The rest of this subsection describes the calculation of the susceptible-normalised log case growth.

In order to smooth out stochastic variations in case reporting, 14 day rolling averages of new case reports, c(t), and cumulative case reports, cc(t), were taken. The population present on any day, n(t), was estimated as described in Section III-C1. A reporting rate, ρ, of $\frac{1}{11}$ was estimated as described in Section IV-A. The susceptible fraction of the population, S, was estimated as in equation 2:

$$S(t) = \frac{n(t) - \frac{1}{\rho} \times cc(t)}{n(t)} \tag{2}$$

This estimate considers cumulative cases as the recovered population ('R' in SIR), with the naive assumption that infected people do not leave the country. This method does not consider the possibility of the same individual being infected twice. However, given that the analysis was restricted to the first 8 months of the pandemic in Andorra, it is highly unlikely that there were significant numbers of reinfections [41].

The susceptible-normalised new cases per day, C, were then estimated as in equation 3:

$$C(t) = \frac{\frac{1}{\rho} \times c(t)}{S(t)} \tag{3}$$

The following analysis uses the log of the growth in daily new cases, LCG, estimated as the relative change in the susceptible-normalised new cases since the 14 days prior (equation 4).

$$LCG(t) = log\left(\frac{C(t)}{C(t-14)}\right) = \log(C(t)) - \log(C(t-14)) \tag{4}$$

Fig. 8 shows time series of the daily LCG(t) values compared to a range of mobility metrics discussed in the next sub-section. Due to many days of zero cases between March and May, the LCG shows some volatility and undefined values.

2) Mobility Metrics: Eight mobility metrics, shown in Fig. 8, were compared to the LCG(t) over time. The mobility metrics all follow a broadly similar trend. They drop off sharply during March and remain close to their minimum levels throughout

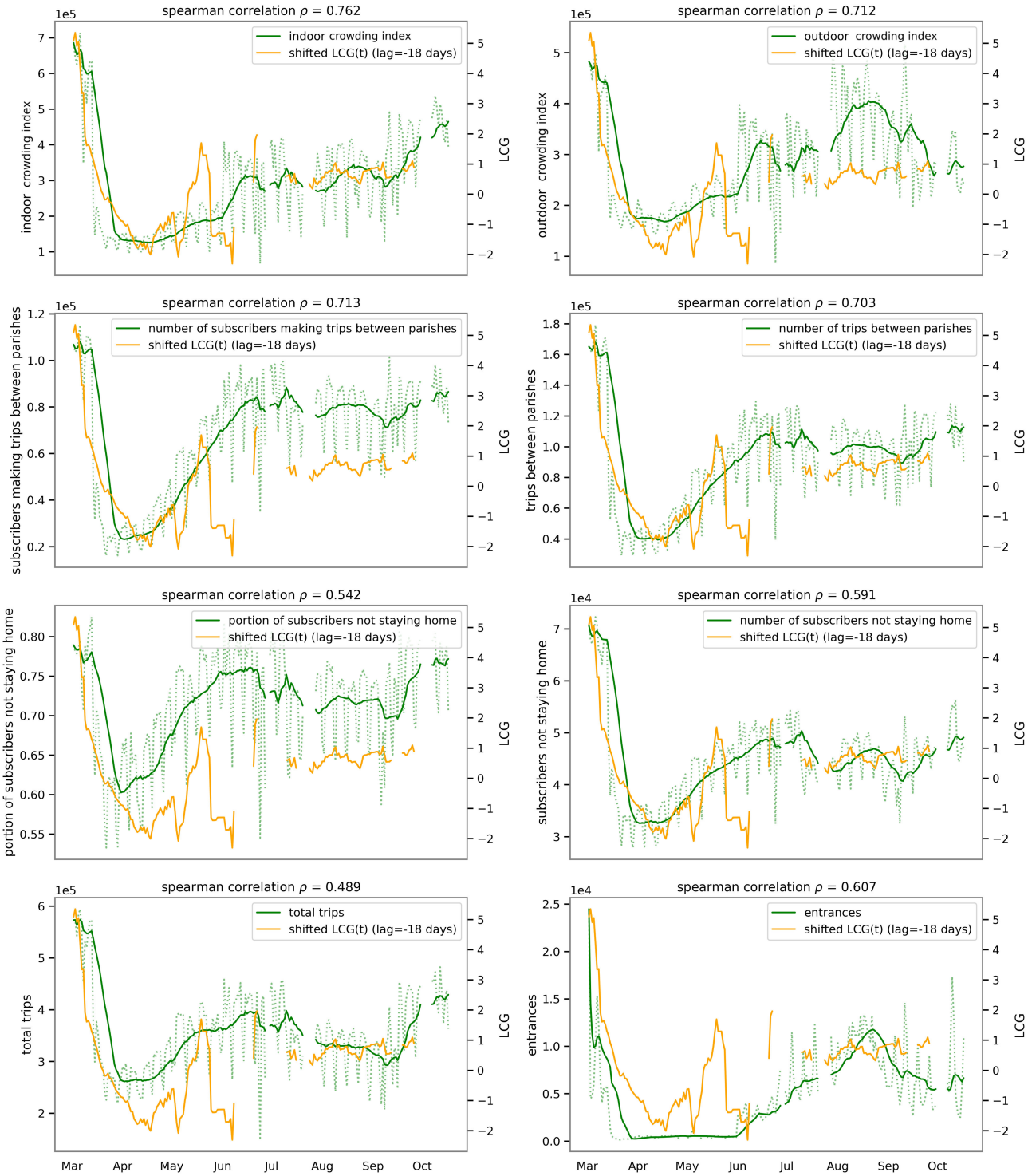

Fig. 8. Time series plots for 8 mobility metrics that were computed from Andorra telecoms data from March - October, 2020. The time series for the estimated log case growth (LCG), shifted by a lag of -18 days is superimposed on the mobility metrics plots to better display the correlations between mobility and transmission rates.

April. During May, the mobility metrics gradually increase, corresponding with a gradual relaxation of policies, or adherence to those policies. The five metrics related to trip making in Fig. 8 ((i) the number of subscribers making trips between parishes, (ii) the number of trips between parishes, (iii) the portion of subscribers not staying home, (iv) the number of subscribers not staying home, (v) total trips) remain relatively steady after June at levels slightly lower than their peak in March. The indoor and outdoor crowding indices in Fig. 8 appear to track one another very closely throughout the period of March to mid-September

but after mid-September, these apparent trends diverge as the indoor crowding index increases while the outdoor crowding index decreases. This is likely due to crowded events moving indoors during the colder months. Entrances to the country, shown in Fig. 8 increase throughout the summer, peaking in late August before decreasing.

3) Correlations between transmission rate and mobility metrics: For any mobility metrics that are correlated with transmission rate, a lag should be expected between changes in these mobility metrics and changes in case growth. If mobility behaviours are responsible for infections, then the mobility metrics would be expected to lead infections by some period. The incubation period for COVID-19 can range up to 14 days [27] and one other study has shown correlations between reproduction rate and a mobility index from Google Mobility data [21] with a 14 day lag [8]. They found an average correlation of 0.63 across multiple regions where these data were available, but did not test other lag periods. However, there may also be further delays between symptom onset, testing and reporting of results. For example, early in the pandemic, COVID-19 testing in Andorra was done by sending samples to labs in foreign countries [16], which may have caused additional days of delay between changes in behaviour and an associated change in reported cases. For these reasons, a broader range of possible lag times are tested for correlations in this study.

On the other hand, previous research has shown that people update their mobility and economic behaviours in response to public reporting of COVID-19 cases [28]. Daily aggregate case numbers for the country were available in Andorra throughout the pandemic. If people were restricting their mobility behaviours in response to changes in the rate of reported infections, then the infections would lead the mobility metrics by some period.

For each metric, it is unknown a priori whether mobility leads infections, infections lead mobility, or both. In order to address this question, the time series data for daily mobility metrics were compared to the daily LCG time series with a range of lag times from −30 to 30 days, where a lag of −30 days corresponds with the mobility metrics leading the LCG time series by 30 days, and a lag of 30 days corresponds with the LCG leading the mobility metrics time series by 30 days. The Spearman's Rank Correlation Coefficient was computed at every lag value. The Spearman's Coefficient was deemed to be more appropriate than the Pearson's Coefficient because it can identify monotonic relationships which are not necessarily linear.

As shown in Fig. 9, several correlation patterns with transmission rate are present across the 8 mobility metrics considered. Fig. 9(a) shows 4 mobility metrics which are highly correlated with transmission rate, leading by 2–3 weeks. This suggests that when the behaviors associated with these mobility metrics change, transmission rates change in a corresponding way about 2–3 weeks later. Conversely, there is no evidence that these metrics increase or decrease in response to changes in transmission rate, as there is a small negative correlation between these mobility metrics and transmission rate when mobility metrics lag LCG. Fig. 9(b) shows 3 metrics which exhibit the same correlation trends as (A), but with slightly weaker correlation values. Again, these metrics are positively correlated with transmission with a lead of 2–3 weeks, and also moderately

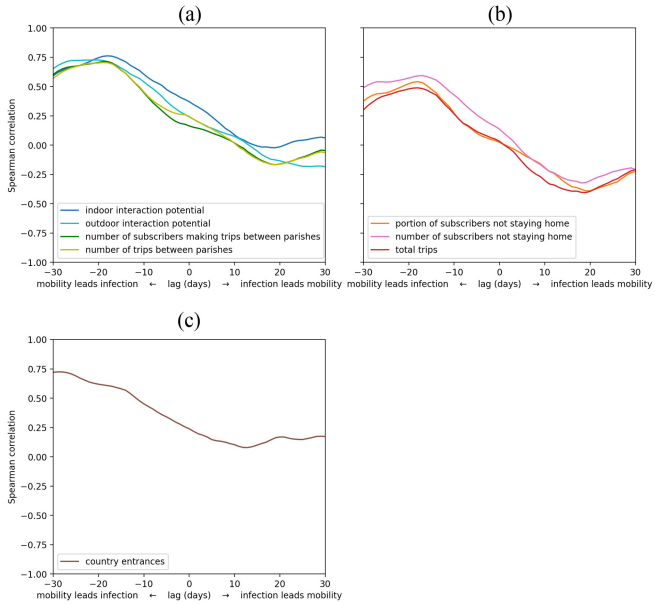

Fig. 9. Daily mobility metrics were computed from the Andorra telecoms data for March–October, 2020. The Spearman correlation between the time series of these metrics and the log case growth (LCG) was computed for lags ranging from −30 to 30 days. Negative lag values correspond with the LCG time series shifted to lag behind the mobility metrics time series. e.g. At a lag of −10 days, mobility at day (t-10) is compared to LCG at day t. Values for 8 mobility metrics are shown in 3 plots. Plot A shows metrics with strong positive correlations with LCG, where changes in mobility lead changes in case growth by 2–3 weeks. Plot B metrics show weaker correlation values at a similar lead time. Plot C shows that the country entrances metric does not have an expected peak in correlation to LCG.

negatively correlated with transmission rate, with a lag of about 2–3 weeks. This suggests that (i) when these mobility behaviours increase, there is a moderate increase in transmission rate 2–3 weeks later and (ii) when the transmission rate increases, there is a moderate decrease in the mobility behaviours 2–3 weeks later. Fig. 9(c) shows that new entrances to the country seem to be highly correlated with infection rate, with a lead of 30 days. However, the correlation remains positive when changes in entrances lag LCG, making the relationship between country entrances and transmission less clear. Lead times greater than 30 days were not explored in this study because relationships between mobility and transmission rate at greater lead times would be less plausible.

V. Discussion

This analysis highlighted a number of insights about the nature of the spread of COVID-19 infection in Andorra. These insights and their potential policy implications are discussed in this section.

Early in the COVID-19 pandemic, some of the most high-profile hotspots were European ski destinations and other areas which experienced high levels of international tourism.

Tourism, especially ski tourism, is a large part of Andorra's economy. This can be shown by the fact that at the start of January 2020, the number of people present in Andorra was about 50% more than Andorra's residential population (see Fig. 5). This quickly transitioned in March when there were

massive departures from the country, as an economic lockdown and border closures were put in effect.

It might have been expected that the areas with highest infection rates in May would be those where the most tourists had visited and interacted with residents before the lockdown. However, in this study it has been shown that there was no correlation between tourist contact and the COVID-19 exposure rates across Andorran parishes as measured in May 2020. In particular, the parish of Canillo accommodated 45% of tourists in February and residents of Canillo had higher crowding indices with respect to tourists than residents of any other parish. Despite this, 3 of the 5 other parishes considered had an equal or higher infection rate. These findings suggest that, while tourists likely imported Andorra's initial COVID-19 cases, subsequent geographical mixing quickly spread the infection to less touristic areas.

More generally, exposures to COVID-19 were distributed across the parishes of Andorra in a way that was unrelated to tourism or measured mobility behaviours. Possible reasons for this are that (i) the spatial resolution of data was coarse (7 parishes), there may have been unobserved socio-demographic differences between parishes which led to differences in infection rate or (iii) there was sufficient mixing across parish populations that the effects of differences in behavior were not long-lasting.

However, changes in the transmission rate over time were closely related to changes in the population's mobility behaviours over time. It was found that trips between parishes were highly correlated with infection rates 2–3 weeks later. This is in contrast to measures of overall trip making (including intra-parish travel), which were only moderately correlated with later infection rates. This may be because trips between parishes can spread the infection between sub-populations at difference stages of the epidemic curve. However trips between parishes may also simply be more indicative of a reopening economy, including the tourism economy which necessitates travel between parishes. The fact that some measures of stay-at-home rates and total trip making were less correlated with transmission rate might suggest that there are safe ways for residents to remain mobile within their communities.

The mobility metrics that were most highly correlated with subsequent transmission rates over time were the crowding indices. Furthermore, this metric was shown to be a fat-tailed phenomenon dominated by a small number of people and events. The crowding indices were calculated based on observed co-locations in H3 cells of resolution 11. However, robustness tests showed that crowding indices calculated at resolutions of 10 and 12 were almost perfectly correlated with this and therefore the choice of resolution would not affect the correlation analysis. Finally, some measures of stay-at-home rates and total trip making could be seen to reduce in response to high infection rates, which may have been due to policy responses or to individual volition.

These findings suggest that, in order to reduce transmission rates, policies should focus on discouraging the most risky behaviours such as dense indoor crowding. Furthermore, the events which draw the highest density of crowds should be identified and discouraged. On the other hand, policies which aim to reduce mobility across the board, are likely to be less effective and have higher social and economic costs. For example, between April 19th and May 13th in Andorra, outdoor exercise was restricted to a 1km radius for 1 hour every second day. Such measures would have likely reduced intra-parish outdoor trips while having little impact on the most relevant metrics like inter-parish trips and crowded indoor events.

The finding that trips between communities were correlated with transmission rate has important implications for simulating the transmission of respiratory diseases since many studies use areas significantly larger than the state of Andorra as the sub-populations. Also, the fact that the crowding indices were highly correlated with transmission rate shows that the density with which people crowd together is an important determinant of transmission rate and therefore, the common assumption of mass-action (as discussed in Section I) is unrealistic for models of human respiratory disease spread.

The combination of population serology testing and telecoms data for all of Andorra provided a uniquely comprehensive dataset to study the progression of the pandemic in an entire country. However, the serology testing only occurred during a single period in May, 2020 and so the analysis of temporal trends in transmission rate were based on the public case report data. The fraction of true cases represented by these reports could not be precisely known. The combination of the serology data and case reports allowed estimation of this reporting factor for the period up to the start of May. However it is likely that the reporting rate subsequently changed with new testing procedures and policies, and as people entered and left the country. The problem of under-reporting is not unique to this study and in many studies there is no accounting for under-reporting as no serology data are available.

VI. Conclusion

During the COVID-19 pandemic, policy makers worldwide were faced with the task of designing policies and interventions to try to curb the spread of infection. Numerous options have been available, including travel bans, economic shutdowns, stay-at-home orders and widespread testing to name a few. Yet all of these options are associated with social and economic costs and unknown levels of effectiveness. This study set out to provide decision support to policy-makers in Andorra and around the world by identifying which types of internal and external mobility are strongly associated with transmission risk and which are not, thereby enabling a greater focus on the most promising policies. A secondary contribution of this work was to provide empirical evidence to inform models for respiratory infection simulations.

The combination of a country-wide serology testing programme and high resolution telecoms data for the entire country presented a unique opportunity for granular insights into the relationship between mobility behaviour and COVID-19 exposure. Additionally, the isolation of Andorra during lockdown and the subsequent reintroduction of tourism provided a natural experiment to examine the relative importance of external versus internal mobility.

It has been shown in this study that internal mobility behaviours such as crowding and rates of inter-community trips were most closely related to the spread of infection in Andorra. Conversely, contact with tourists, intra-community trips and

entrances to the country had lower or unclear correlations with transmission rates. Furthermore, crowding was characterised by highly skewed distributions with small numbers of people and events accounting for large proportions of the human contact.

Overall, the results of this study point policy-makers towards a targeted approach to mobility restrictions to curb the spread of COVID-19, based on discouraging the most risky behaviours such as gathering in crowed indoor settings.

References

[1] *Overview of the h3 Geospatial Indexing System*. Accessed: Dec. 22, 2020. [Online]. Available: https://h3geo.org/docs/core-library/overview

[2] *Ski, Party, Seed a Pandemic: The Travel Rules That Let COVID-19 Take Flight*. Accessed: Dec. 22, 2020. [Online]. Available: https://www.nytimes.com/2020/09/30/world/europe/ski-party-pandemic-travel-coronavirus.html

[3] *A Ski Resort Spread Coronavirus Across Europe. Austria and Switzerland Are Eager to Reopen Slopes Anyway*. Accessed: Dec. 22, 2020. [Online]. Available: https://www.washingtonpost.com/world/2020/11/26/european-ski-resorts-reopen-despite-coronavirus/

[4] *Table of Cell Areas for h3 Resolutions*. Accessed: Dec. 22, 2020. [Online]. Available: https://h3geo.org/docs/core-library/restable

[5] A. Aleta *et al.*, "Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic," *Nat. Hum. Behav*, vol. 4, pp. 964–971, 2020.

[6] F. E. Alvarez, D. Argente, and F. Lippi, *A Simple Planning Problem for COVID-19 Lockdown*. Working Paper 26981, National Bureau of Economic Research, Apr. 2020.

[7] R. M. Anderson, and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. London, U.K: Oxford Univ. Press, 1992.

[8] F. Arroyo-Marioli, F. Bullano, S. Kucinskas, and C. Rondón-Moreno, "Tracking R of COVID-19: A new real-time estimation using the Kalman filter," *PLoS One*, vol. 16, no. 1, 2021, Art. no. e0244474.

[9] M. Bakker, A. Berke, M. Groh, A. Pentland, and E. Moro, *Effect of Social Distancing Measures in the New York City Metropolitan carea*. Boston, MA, USA: Massachusetts Inst. Technol., 2020.

[10] J. M. Brauner *et al.*, "Inferring the effectiveness of government interventions against COVID-19," *Science*, vol. 371, no. 6531, 2021.

[11] S. Chang *et al.*, "Mobility network models of COVID-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 1–6, 2020.

[12] CNN'S Pierre Meilhan - CNN France closes its borders to contain the coronavirus outbreak. Accessed: Feb. 4, 2021. [Online]. Available: https://edition.cnn.com/world/live-news/coronavirus-outbreak-03-16-20-intl-hnk/h_59fc12079ee007cb3db50e20cb26722e

[13] C. Corbane, P. Politis, V. Syrris, and M. Pesaresi, "GHS built-up grid, derived from Sentinel-1 (2016), R2018A," *Eur. Commission, Joint Res. Centre*. Accessed: Sep. 10, 2018. [Online]. Available: http://data.europa.eu/89h/jrc-ghsl-1000

[14] C. L. Correa-Martínez *et al.*, "Pandemic in times of global tourism: Superspreading and exportation of COVID-19 cases from a ski area in Austria," *J. Clin. Microbiol.*, vol. 58, no. 6, 2020, Art. no. e00588-20.

[15] Departament-d'Estadística. Estimacions de població, gener 2020. A003. Estadstica dels censos parroquials, gener 2020, pp. 1–11, Accessed: Aug. 2021. [Online]. Available: https://www.estadistica.ad/serveiestudis/noticies/noticia5059cat.pdf

[16] Diari D'andorra. Mor un home de 88 anys amb coronavirus., 2020. Accessed: Aug. 2021. [Online]. Available: https://www.diariandorra.ad/noticies/nacional/2020/03/22/mor_home_anys_amb_coronavirus_158881_1125.html

[17] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, 2020.

[18] C. Corbane, P. Politis, V. Syrris, and M. Pesaresi, "GHS built-up grid, derived from Sentinel-1 (2016), R2018A," 2018, Accessed: Aug. 2021. [Online]. Available: https://data.jrc.ec.europa.eu/dataset/jrc-ghsl-10008

[19] Facebook. Facebook data for good. 2020. Accessed: Jul. 2020. [Online]. Available: https://dataforgood.fb.com/docs/covid19/

[20] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, 2008, Art. no. 779.

[21] Google. Google COVID-19 community mobility reports. 2020. Accessed: Jul. 2020. [Online]. Available: https://www.google.com/covid19/mobility/

[22] Govern D'andorra. Els tests d'anticossos permeten diagnosticar 78 positius de la COVID-19, que podrien haver contagiat unes 360 persones, 2020. Accessed: Aug. 2021. [Online]. Available: https://www.govern.ad/comunicats/item/11629-els-tests-d-anticossos-permeten-diagnosticar-78-positius-de-la-covid-19-que-podrien-haver-contagiat-unes-360-persones

[23] M.-G. Hâncean, M. Slavinec, and M. Perc, "The impact of human mobility networks on the global spread of COVID-19," *J. Complex Netw.*, vol. 8, no. 6, 2020, Art. no. cnaa041.

[24] N. Haug *et al.*, "Ranking the effectiveness of worldwide COVID-19 government interventions," *Nature Human Behav.*, vol. 4, no. 12, pp. 1–10, 2020.

[25] R. F. Hunter *et al.*, "Effect of COVID-19 response policies on walking behavior in U.S. cities," *Nature Commun.*, vol. 12, no. 1, pp. 1–9, 2020.

[26] M. J. Keeling, and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ, USA: Princeton Univ. Press, 2011.

[27] S. A. Lauer *et al.*, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application," *Ann. Intern. Med.*, vol. 172, no. 9, pp. 577–582, 2020.

[28] K. O. Lee, and H. Lee, "Public responses to COVID-19 case disclosure and their spatial implications," *SSRN*, 2020, Accessed: Aug. 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3758394

[29] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, "Mining user similarity based on location history," in *Proc. 16th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2008, pp. 1–10.

[30] Livzon Diagnostics. The Diagnostic Kit for IgM/IgG Antibody to Coronavirus (SARS-CoV-2) (Lateral Flow). Accessed: Feb. 4, 2021. [Online]. Available: http://www.livzondiagnostics.com/en-us/info/17.html

[31] B. J. McNeil, and S. J. Adelstein, "Determining the value of diagnostic and screening tests," *J. Nucl. Med.*, vol. 17, no. 6, pp. 439–448, 1976.

[32] Ministerio de asuntos exteriores, Unió n europea y cooperación. Frontera España-Andorra ante las crisis del COVID 19. Accessed: Feb. 4, 2021. [Online]. Available: http://www.exteriores.gob.es/Embajadas/Andorra/es/Noticias/Paginas/Articulos/19032020_NOT01.aspx

[33] A. Noyman, R. Doorley, Z. Xiong, L. Alonso, A. Grignard, and K. Larson, "Reversed urbanism: Inferring urban performance through behavioral patterns in temporal telecom data," *Environ. Plan. B: Urban Analytics City Sci.*, vol. 46, no. 8, pp. 1480–1498, 2019.

[34] E. Pepe *et al.*, "COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown," *Scientific Data*, vol. 7, no. 1, pp. 1–7, 2020.

[35] S. Phithakkitnukoon, Z. Smoreda, and P. Olivier, "Socio-geography of human mobility: A study using longitudinal mobile phone data," *PLoS One*, vol. 7, no. 6, 2012, Art. no. e39253.

[36] H. V. Ribeiro, A. S. Sunahara, J. Sutton, M. Perc., and Q. S. Hanley, "City size and the spreading of COVID-19 in Brazil," *PLoS One*, vol. 15, no. 9, 2020, Art. no. e0239699.

[37] W. J. Rogan, and B. Gladen, "Estimating prevalence from the results of a screening test," *Amer. J. Epidemiol.*, vol. 107, no. 1, pp. 71–76, 1978.

[38] C. Royo-Cebrecos *et al.*, "Mass SARS-CoV-2 serological screening, a population-based study in the principality of Andorra," *Lancet Regional Health-Europe*, Elsevier, vol. 5, 2021, Art. no. 100119.

[39] Safegraph. Social distancing metrics, 2020. Accessed: Jul. 2020. [Online]. Available: https://www.safegraph.com/data-examples/covid19-shelter-in-place/

[40] M. J. Siedner *et al.*, "Social distancing to slow the us COVID-19 epidemic: Longitudinal pretest-posttest comparison group study," *PLoS Med.*, vol. 17, no. 8, 2020, Art. no. e1003244.

[41] C. Stokel-Walker, "What we know about COVID-19 reinfection so far," *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n99.

[42] M. S. Warren, and S. W. Skillman, "Mobility changes in response to COVID-19," 2020, *arXiv:2003.14228*.

[43] Y. Zhou, R. Xu, D. Hu, Y. Yue, Q. Li, and J. Xia, "Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: A modelling study using mobile phone data," *Lancet Digit. Health*, vol. 2, no. 8, pp. e417–e424, 2020.