

1 **Size and duration of COVID-19 clusters go along with a high SARS-**
2 **CoV-2 viral load : a spatio-temporal investigation in Vaud state,**
3 **Switzerland**

4
5 Anaïs Ladoy^{1,7}, Onya Opota², Pierre-Nicolas Carron³, Idris Guessous^{4,7}, Séverine
6 Vuilleumier^{5♦}, Stéphane Joost^{1, 5, 6, 7♦}, Gilbert Greub^{2, 8♦†}

7
8 ¹Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental
9 Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

10 ²Institute of Microbiology, University Hospital Centre and University of Lausanne, Switzerland

11 ³Department of Emergency Medicine, Lausanne University Hospital, Rue du Bugnon 46, CH-1011, Lausanne,
12 Switzerland

13 ⁴Division and Department of Primary Care Medicine, Geneva University Hospitals, Geneva, Switzerland

14 ⁵La Source School of Nursing, University of Applied Sciences and Arts Western Switzerland (HES-SO),
15 Lausanne, Switzerland

16 ⁶Unit of Population Epidemiology, Division of Primary Care, Geneva University Hospitals, Switzerland

17 ⁷Group of Geographic Information Research and Analysis in Population Health (GIRAPH), Switzerland

18 ⁸ Infectious Diseases Service, University Hospital Centre, Lausanne, Switzerland

19

20 ♦ Equally contributed

21 † Corresponding author¹

22

23 **Running title**

24 Spatio-temporal evolution of COVID-19 clusters

25

26

27

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

¹ Gilbert.greub@chuv.ch

28 **Abstract**

29 To understand the geographical and temporal spread of SARS-CoV-2 during the first wave of
30 infection documented in the canton of Vaud, Switzerland, we analysed clusters of positive
31 cases using the precise place of residence of 33'651 individuals tested (RT-PCR) between
32 January 10 and June 30, 2020. We identified both space-time (SaTScan) and transmission
33 (MST-DBSCAN) clusters; we estimated their duration, their transmission behavior
34 (emergence, growth, reduction, etc.) and relative risk. For each cluster, we computed the within
35 number of individuals, their median age and viral load.

36 Among 1'684 space-time clusters identified, 457 (27.1%) were significant ($p \leq 0.05$), i.e.
37 harboring a higher relative risk of infection, as compared to other regions. They lasted a median
38 of 11 days (IQR 7-13) and included a median of 12 individuals per cluster (IQR 5-20). The
39 majority of significant clusters ($n=260$; 56.9 %) had at least one person with an extremely high
40 viral load (above 1 billion copies/ml). Those clusters were considerably larger (median of 17
41 infected individuals, $p < 0.001$) than clusters with subjects showing a viral load lower than 1
42 million copies/ml (median of 3 infected individuals). The highest viral loads were found in
43 clusters with the lowest average age, while clusters with the highest average age had low to
44 middle viral load. Interestingly, in 20 significant clusters the viral load of three first cases were
45 all below 100'000 copies/ml suggesting that subjects with less than 100'000 copies/ml may still
46 have been contagious. Noteworthy, the dynamics of transmission clusters made it possible to
47 identify three diffusion zones, which mainly differentiated rural from urban areas, the latter
48 being more prone to last and spread in a new nearby clusters.

49 The use of geographic information is key for public health decision makers to mitigate the
50 spread of the virus. This study suggests that early localization of clusters help implementing
51 targeted protective measures limiting the spread of the SARS-CoV-2 virus.

52

53 **Introduction**

54 The novel coronavirus SARS-CoV-2 causing the COVID-19 disease has impacted our
55 societies on an unprecedented scale. The number of infected people increased rapidly

56 globally, with more than 84 million confirmed cases as of January 2021 and more than 1.8
57 million deaths (1, 2). The quick spread of the disease has challenged international experts and
58 policymakers to implement strategies according to the local virus spread, healthcare
59 resources, economic and political factors (Nicola et al., 2020; see also the cross-country
60 analysis of COVID-19 response: <https://analysis.covid19healthsystem.org/>). Around the globe,
61 tracing, lockdowns and quarantines have been implemented to contain the spread and
62 impacted more than four billion people worldwide (4). These measures aim to protect the
63 significant fraction (~22%) of the world population at risk of severe COVID-19 (5). They raise
64 major challenges related to their dramatic impact on individual health, the capacities of our
65 healthcare system and on our economy (4, 6, 7).

66 COVID-19 outbreaks occur by spreading via close contact forming clusters of cases. A critical
67 challenge to contain the spread of the virus lies (i) in the early detection of these clusters, which
68 reflects active viral transmission (8) and (ii) in the understanding of their spatial and temporal
69 evolution (9). Geospatial tools using the precise location of the place of residence of tested
70 individuals are highly effective to oversee an epidemic (10, 11). They allow for implementation
71 of strategies to control the local disease spread in space and time (12).

72 Although widely used, there is no general agreement on the definition and the concepts of
73 cluster, outbreak and hotspot, and more specifically in a spatial context. The information
74 available from public health departments around the world globally converge, despite
75 differences exist. The term “cluster” generally refers to a temporal aggregation and a spatial
76 concentration of infections cases. COVID-19 clusters are constituted of two or more test-
77 confirmed cases – three or more in France (www.santepubliquefrance.fr) and Switzerland,
78 even 10 or more in New Zealand (www.health.govt.nz) – among individuals associated with a
79 specific non-residential setting with illness onset dates within 7 to 14 days. To further label
80 clusters as outbreak, one must also either 1) have identified direct exposure between at least
81 two of the test-confirmed cases in that setting (for example under one meter face to face) and
82 this during the infectious period of one of the cases, or 2) when there is no sustained local
83 community transmission, one must have noticed the absence of an alternative source of

84 infection outside the setting for the initially identified positive cases (13). Clusters are also
85 assimilated to the concept of “hotspot”, which is not clearly defined neither but often used in
86 spatial epidemiology (14). The World Health Organization (WHO) defined a set of methods
87 and procedures to identify epidemic hotspots and to use them for global surveillance of a
88 population (UNAIDS/WHO, 2013). Beside, infectious diseases studies have proposed
89 methods to identify spatial clusters and characterize them (16, 17).

90 The identification of areas of high prevalence of any phenomenon constitutes a specific
91 research domain in spatial statistics. Point pattern analysis (18) and local spatial
92 autocorrelation methods were applied for the detection of disease clusters (19). In the current
93 COVID-19 pandemic, Zhang et al. (2020) used local Moran’s statistics to identify clusters in
94 China, but at a large geographic scale and using incident cases aggregated at the level of
95 large administrative units. Among studies involving geospatial information reviewed by Franch-
96 Pardo et al. (2020), few characterized the spread of COVID-19 in space and time (e.g.
97 Desjardins et al., 2020) and even less used spatial statistics to detect clusters at a local scale
98 (22, 8). More studies on local and regional scales that consider demographic characteristics
99 of a population at risk are needed to provide timely information to enable accurate prevention
100 and containment measures (10). Indeed, the precise detection of spatial clusters, the
101 description of their dynamics and evolution over time in a geographic context are key to inform
102 decision-makers, to deploy smart testing overtime and to provide targeted health and
103 prevention interventions at a local scale (23).

104 The persistence in time of clusters was shown to be associated with socio-economic
105 deprivation (22), but their size and duration are also likely to be due to so-called “super-
106 spreader” individuals or events (24). They relate to the evidence for large variation in individual
107 reproductive number (25). A super-spreader is considered to greatly contribute to the
108 transmission of an infectious disease. Stein (2011) estimates that it would correspond to a
109 20/80, i.e. 20% of individuals for up to 80% of the transmission. Super-spreaders exist among
110 the SARS-CoV-2 infected persons (27); they are more likely to be highly infectious, a
111 mechanism suggested to be related to high viral loads (28). Noteworthy, as shown recently

112 (29), viral load found in SARS-CoV-2 infected people appears to be similar to that observed
113 with other respiratory viruses such as influenza B and to be similar across ages. Why SARS-
114 CoV-2 exhibits such a high reproductive number (R_0) of about 2 to 3.5 (30), and whether
115 transmission pattern, cluster duration and size correlate somehow with viral load remains to
116 be explored in a detailed spatio-temporal context.

117 Here, we characterized the spatial and temporal dynamics of the first wave of SARS-CoV-2
118 infection in the canton of Vaud (western Switzerland) through the detection and the location of
119 clusters, as well as of their characteristics such as size, duration and composition (number and
120 age of individuals and their viral load). We used the results of the SARS-CoV-2 RT-PCR tests
121 ($n= 33'651$) performed by the Microbiology Laboratory of the Lausanne University Hospital
122 (CHUV) between January 10 and June 30, 2020 (with a first positive case on 2nd March). The
123 data collected are results of RT-PCR tests, viral load (copies/ml) when the test is positive, age
124 and geographic location of the address of residence of individuals tested. We used on the one
125 hand a spatial scan approach (31, 32) (i) to detect spatio-temporal clusters of COVID-19 on a
126 daily basis, (ii) to disentangle the relationships between cluster size, duration and composition,
127 and (iii) to assess the importance of viral load in the evolution of the clusters. On the other
128 hand, we implemented the Modified Space-Time DBSCAN (MST-DBSCAN) algorithm (33) to
129 characterize the diffusion dynamics of transmission clusters. Finally, we discussed the effect
130 of a soft lockdown such as deployed in Switzerland from March 19 to April 27, 2020, on the
131 dynamics of the spread of the virus.

132

133 **Material and Methods**

134 *Patients*

135 Patients exhibiting symptoms compatible with COVID-19, such as fever, cough, dyspnea,
136 smell loss or taste loss were tested by RT-PCR for the presence of the SARS-CoV-2 in their
137 nasopharyngeal secretions, at least when considered vulnerable (e.g. with
138 immunosuppression, obesity, chronic obstructive lung disease or age > 65 years) or when
139 likely exposed to vulnerable cases (e.g. healthcare workers or subject living with vulnerable

140 persons). Contacts of most positive cases were also tested, even when asymptomatic, in order
141 to define need of a 10 days quarantine period or isolation. Precise address of residence was
142 prospectively collected at the time of sampling as well as person age.

143

144 *SARS-COV-2 RT-PCR*

145 Most RT-PCR were performed using the automated molecular platform implemented at the
146 Institute of Microbiology. It uses the Magnapure automated RNA extraction method followed
147 by PCR amplification on QuantStudio automated systems (34) with primers described by
148 Corman et al. (2020), later slightly modified according to Pillonel et al. (2020) to further improve
149 PCR sensitivity. Then, from March 24, 2020, most RT-PCR were performed using the COBAS
150 6800 RT-PCR test, which exhibited similar performance than the home-brew automated
151 approach (Opota et al. 2020). A few numbers of cases were tested using the GeneXpert
152 approach to reduce time to results (38). Viral load was calculated based on the so-called “cycle
153 threshold” (Ct), which corresponds to the number of cycles, when the fluorescent signal is
154 above a predefined threshold (37, 38).

155

156 *Study area*

157 All the data used were collected in the state of Vaud located in the south-west of Switzerland,
158 north of Lake Geneva. It has an area of 3'212km² (see Figure 6A), a population of 811'203
159 individuals (end of 2019), for a density of 249 inhabitants/km². Of note, there are important
160 differences in population density between the urban area of Lausanne-Morges on the shores
161 of Lake Geneva (~3'000 inhabitants/km²), and country-side toward north where population
162 density is of ~200 inhabitants/km². One exception is the area of Yverdon-les-Bains directly
163 south of the Lake of Neuchâtel with 2'200 inhabitants/km².

164

165

166

167 *Spatio-temporal clusters*

168 We used the SaTScan software (version 9.6.1) to detect daily space-time clusters of
169 individuals tested positive for SARS-COV-2 in the canton of Vaud from March 2 to June 30
170 2020 (no positive cases between January 10 and March 2, 2020). The algorithm developed by
171 Kulldorff (1997) tests whether a disease is randomly distributed over space and time. It uses
172 a “moving cylinder”, with the base and height corresponding to the spatial and temporal
173 components, respectively. Significance evaluates the excess relative risk, i.e. more than
174 expected observed COVID-19 cases within the moving cylinder relative to randomly distributed
175 cases over space and time. We implemented it in a daily prospective surveillance analysis.
176 We used a discrete Poisson model, where the number of events in the geographic area (total
177 number of positive tests) is Poisson-distributed, according to a known underlying population at
178 risk. The spatial size of the clusters’ radius reported on maps covers a maximum of 0.5% of
179 the total resident population (population at risk) in the canton of Vaud (N=811’203 inhabitants;
180 SFSO, 2019). Tested individuals and the underlying population at risk were georeferenced at
181 the centroids of a hectometric grid (40) covering the entire study area. The minimum number
182 of positive cases considered to constitute a cluster is 3, and we restricted the temporal
183 scanning window to a minimum of 2 days and a maximum of 14 days. The upper limit of 14
184 days accounts for the incubation period (generally 2 to 7 days) and for being infectious (no
185 more than 10 days from symptoms onset). The significance of the clusters was evaluated on
186 the basis of 999 Monte-Carlo permutations that randomize both locations (41) and times of the
187 cases.

188

189 *Cluster evolution and diffusion zones*

190 We used MST-DBSCAN (modified space–time density-based spatial clustering of application
191 with noise; Kuo et al., 2018) to characterize the diffusion dynamics of clusters. MST-DBSCAN
192 is an algorithm for detecting, characterizing, and visualizing disease cluster evolution in the
193 geographic space and in time. It computes geographically a kernel density that considers the
194 effect of the incubation period of an infection disease. It is based on DBSCAN (42), a non-

195 parametric density-based clustering algorithm that groups together objects (here SARS-COV-
196 2 positive cases) that are closely packed together (points with many nearby neighbors),
197 marking as outliers points that lie in low-density regions. The MST-DBSCAN identifies six
198 different cluster behaviors: a) emerge, b) grow (or increase), c) remain steady (keep), d) move,
199 e) split or f) reduce (decrease).

200 We applied the MST-DBSCAN analysis to the 3'317 COVID-19 positive cases identified
201 (among 33'651 tested individuals) and georeferenced at their precise address of residence in
202 the canton of Vaud. Disease clusters were computed daily from March 4, 2020, to June 30,
203 2020. The maximum spatial radius considered was of 1'000 meters, and we have set a time
204 window of 1 to 7 days to reflect the incubation period of the disease. A minimum number of 3
205 positive cases was considered to constitute a cluster. For all clusters identified, we established
206 a typology of similar diffusion patterns in the geographical space. We associated the clusters
207 with the postcode areas (557 units; MicroGIS, 2019) of the canton of Vaud and used them as
208 spatial references. Then, we focused on three main cluster behaviors, which are increase (b),
209 reduce (f) and keep (c) to characterize the diffusion type through the postcode areas. The
210 diffusion patterns were detected using the Louvain method, a group detection algorithm using
211 network analysis (44). This approach synthetizes the spatio-temporal information and
212 facilitates its visualization on a single map.

213

214 **Results**

215 *Epidemic trajectories of positive cases*

216 A total of 33'651 subjects have been tested over a period of 6 months, of which 3'317 (9.86%)
217 were positive by RT-PCR. Seventy-nine percent of positive cases (2609/3317) were observed
218 between March 9 and April 5: this 4-weeks period corresponds to only 16% of the total duration
219 of the studied period, but in this short period as many as 2'609 tests were positive (22.2% of
220 11'756 tests) (Figure 1A). The peak of the 1st epidemic wave occurred on March 18, i.e. two
221 days after the start of the soft lockdown implemented in Switzerland and lasting from March
222 16 to April 27 (see vertical dashed lines in Figure 1A). At the peak of the outbreak, as many

223 as 180 Vaud subjects were documented positive in our laboratory in a single day (Figure 1A,
224 dark blue). Number of positive cases decreased then considerably from May 1st. The highest
225 proportion of positive tests was observed 4 days after the peak of the epidemic wave, with a
226 rate of positive tests reaching 32% (Figure 1A, light blue). The rate of positive cases was
227 relatively high at the start of the epidemic when few individuals were tested and found positive.
228 Then the shape of percentage of positive tests followed the trajectory of the number of cases.
229 This is likely to be related to the fact that at the beginning of the epidemics only symptomatic
230 subjects and patient at risk were tested, and then a much wider category of individuals was
231 tested to the point all symptomatic individuals and asymptomatic contacts could access a test.
232

233 *Cluster detection and temporal dynamics*

234 We identified 1'684 space-time clusters using the place of residence of patients positive to
235 SARS-CoV-2. Among them, 457 were considered significant based on the within proportion of
236 positive cases compared to the total documented positive cases. Highest values of both
237 significant and non-significant clusters were observed between March 9 and April 5 (Figure
238 1B). Number of clusters decreased from the 1st of May. Thus, the decrease in the number of
239 positive patients following the beginning of the soft lockdown (Figure 1A) occurred about two
240 weeks earlier than the decrease in the number of clusters (Figure 1B). Significant or not, the
241 number of clusters displays a similar pattern through time in terms of increase and decrease
242 but with a difference in amplitude. As shown on Figure 1C, the relative risk for new clusters
243 was higher before the soft lockdown and about 80 days after the end of the lockdown. The size
244 of clusters (i.e. number of cases within clusters) used to compute the relative risk does not
245 strongly change the value of the relative risk during the core of the epidemic wave; however,
246 it affects this value when the number of positive cases is small, i.e. at the beginning and at the
247 end of the epidemic wave.

248

249

250

251 *Cluster composition*

252 Significant space-time clusters generally involved a larger number of positive cases (maximum
253 of 21 cases in average, with the largest cluster showing 43 cases on March 25) compared to
254 non-significant ones (maximum of 11 positive cases in average; Figure 2A). Noteworthy, as
255 shown in Figure 2A, significant clusters with more than 15 positive cases were mainly observed
256 shortly after the soft lockdown implemented from March 16 to April 27, with one exception on
257 April 3. Cluster durations – although limited to 14 days - increased over time from the start of
258 the epidemic wave, showing little differences between significant and non-significant clusters.
259 We can also observe the absence of any significant cluster from May 3 to June 16 (Figure 2B).

260

261 *Viral load in clusters*

262 Clusters were defined by the presence of at least 3 positive cases within a limited geographic
263 area, as documented in the SaTScan analysis. All clusters were then characterized according
264 to the nasopharyngeal viral load of the cases documented in each cluster (Table 1). Significant
265 clusters can have patients with viral load as low as those found in non-significant clusters, i.e.
266 even below 10'000 copies/ml (Supp. Mat. 3). Thus, 5 significant clusters were composed of 3
267 cases exhibiting a viral load below 10'000 copies/ml at time of testing. However, significant
268 clusters were more likely to be detected when viral loads were above 100 million copies/ml
269 (Figure 3). Finally, as many as 18 significant clusters with at least 1 subject showing between
270 1 billion and 10 billion copies/ml were documented on March 24 (Figure 3, pink curve). The
271 frequency distribution of viral load in significant clusters significantly differs from the distribution
272 of viral load in non-significant clusters and outside clusters (Kolmogorov–Smirnov test, two-
273 sample case, $p < 0.001$, see Supp. figures 1 and 2).

274 The mean viral load of the first 3 cases was also studied, in order to gain insight of the possible
275 relationship between nasopharyngeal viral load and contagiousness, indirectly measured by
276 the documentation of subsequent clusters. For 20 significant clusters, all first 3 cases exhibited
277 a viral load below 100'000 copies/ml, suggesting that subjects with less than 100'000 copies/ml

278 may still be contagious (Supp. Mat. 4). Moreover, the nasopharyngeal viral load of the first 3
279 cases was below 1 million copies for 40 significant clusters.

280

281 *Cluster size, duration and viral load*

282 Cluster size (number of within cases) is positively associated with the presence of individuals
283 with high viral load (Figure 4A). The highest viral loads measured showed a value of at least
284 10 billion copies/ml and occurred in the largest clusters (median number of 21 positive cases).
285 Such a result nicely identifies super-spreading events. Even when comparing clusters
286 harboring individuals with all viral loads below 1 million copies/ml and the ones with at least
287 one case showing a viral load above 1 million copies/ml, the difference in cluster size was
288 significant with a median increasing from 3 cases per cluster to 4 cases per cluster ($p < 0.001$;
289 figure 4A). Similar relationships were observed when considering the mean and maximal
290 values of viral loads of the first three positive cases (Figure 5A & B).

291 Highest values of viral load were found in clusters with individuals showing the lowest average
292 age. Clusters composed of individuals with the highest average age had low to middle viral
293 load values (Figure 4B). Indeed, the median age of the individuals within a cluster is
294 significantly higher when the viral load value in the cluster is between 1 and 10 million
295 copies/ml. Then, the average age progressively decreases from 74 years to 48 years, while
296 viral load values increase. Cluster duration is significantly different only between the orange
297 category (100 million to 1 billion/ml) and the pink category (1 to 10 billion/ml) as clusters of the
298 latter category last longer (mean of +0.456 days, $p < 0.001$; see Figure 4C).

299 Interestingly, clusters with individuals showing the lowest average age and the highest viral
300 load (Figure 4B) also constitute the largest clusters (Figure 4A) and those that last the longest
301 (Figure 4C).

302

303 *Geographic distribution of the first epidemic wave*

304 We chose six key dates to illustrate the evolution of the two types of clusters during the first
305 wave of the epidemic in the canton of Vaud. An animation showing the spatio-temporal

306 evolution of the clusters for the whole first wave can be visualized in Supp. Mat. 5, and in Supp.
307 Mat. 6 for the dynamics of clusters' behavior. Here, Figure 6 shows the spatial distribution of
308 space-time clusters (A-F) and compares it to information translating the diffusion dynamics of
309 the clusters (A'-F'). A detailed description of the first SARS-CoV-2 epidemic wave in the state
310 of Vaud can be found in Box 1, illustrating the powerful and critical information that the
311 approach offers.

312 Cluster behaviors described in Box 1 were summarized with four diffusion zones shown in
313 Figure 7A and identified at the level of postal code areas using MST-DBSCAN (Figure 7BCD).
314 The grey diffusion zone corresponds to areas where no clusters emerged, while the green,
315 orange and blue diffusion zones differ in the way clusters evolved over time. The green
316 diffusion zones correspond to areas where the clusters immediately increased in size at the
317 beginning of the epidemic wave (red line, Figure 7B) but decreased drastically once the soft
318 lockdown (vertical dash line) took place. Then we observed a second peak associated with an
319 important increase of clusters that reduced in size (red line, Figure 7B). Both red and purple
320 curves are bimodal and tend to decrease afterwards, with a few numbers of new small peaks
321 that plateau forming a distribution with a long right tail. Conversely, orange and blue diffusion
322 zones show a first peak of increasing clusters later, i.e. about at the time of the start of the soft
323 lockdown (orange & blue areas, Figure 7C and D). Both also show clusters that remain stable
324 in size during the soft lockdown (blue line). Blue diffusion zone is the only one to show no more
325 clusters after April 27, i.e. the end of the lockdown and that does not display a bimodal
326 distribution of clusters. Please note that no difference in viral load was documented among
327 these different diffusion zones (Figure 7E).

328

329 **Discussion**

330 The discussion is divided into three major parts. The first highlights results that uncover new
331 information on COVID-19 clusters, the second explicits limitations in the interpretation of the
332 results, the third presents added value of the methods used to tackle epidemics and to evaluate
333 the effect of lockdown strategies.

334 **New information on COVID-19 clusters**

335 ***A temporal lag between positive cases documentation and clusters burden***

336 Significant clusters were mainly observed from March 15 to April 5 (red curve in Figure 1B),
337 whereas non-significant clusters that occurred specifically in high population density areas
338 such as Lausanne were already documented 4 to 5 days earlier and continuously occurred
339 until mid-May (grey curve on Figure 1B, and Figure 6A). We observed a time shift between the
340 decrease of the number of positive cases and the decrease of the number of clusters. This
341 delay could be explained by the fact that most positive cases might have been at the origin of
342 lasting clusters, i.e. lasting more than 10 days starting at the time when positive cases are
343 identified. Interestingly, the number of patients hospitalized at Lausanne University Hospital
344 (CHUV) and the number of deaths due to COVID-19 in the Canton of Vaud also followed the
345 same epidemic curve, but with a 2 weeks delay (personal communication, G. Greub).

346

347 ***Viral load is strongly informative on the presence and size of SARS-CoV-2 clusters***

348 Our results show that clusters at the peak of the SARS-CoV-2 epidemic wave are composed
349 of individuals showing a high viral load. Cluster size is positively associated with the presence
350 of individuals with a high viral load among the significant clusters, although 40 clusters had
351 their 3 first cases exhibiting a viral load below 1 million copies/ml, including 33 clusters that all
352 had their cases with a nasopharyngeal viral load below 1 million copies. Moreover, as many
353 as 20 clusters were composed of cases that initially all had a viral load below 100'000
354 copies/ml, suggesting that subjects with less than 100'000 copies/ml may still have been
355 contagious.

356 The fact that significant clusters are composed of patients with viral load as low as those found
357 in non-significant clusters further supports the hypothesis of community transmission with low
358 level of viral load. Nevertheless, this may also reflect a statistic bias since large clusters with
359 more than 10 individuals are more likely to get at least one individual with a very high viral
360 load.

361

362 ***Advantage of RT-PCRs over antigen-based testing***

363 Given the relatively low sensitivity of antigen tests, if we had used these assays, about 24
364 clusters would have been missed or identified late, when the first 3 cases would have led to
365 additional cases. Indeed, the 20 significant clusters with a viral load of the three first cases
366 below 100'000 copies/ml would not have been detected with antigen tests, given their limit of
367 detection (for the best ones) of about 100 to 200'0000 copies/MI (45). Moreover, the clusters
368 with a case between 100'000 copies/ml and 1 million copies/ml will also not have been
369 detected in 5% of cases given an overall antigen sensitivity for such viral load of about 80%
370 (45). Thus, altogether, an antigen-based strategy would miss about 5% (24/457) of the
371 significant clusters.

372

373 ***High viral load in large clusters with the youngest group age***

374 Within clusters we found a clear negative relationship between age and level of viral load
375 measured (Figure 4B), and between cluster size and viral load (Figure 4A). Indeed, while a
376 high viral load was found in large clusters with the youngest group age, low to intermediate
377 viral load was measured in small clusters constituted of older group age. This suggests that
378 large clusters were generated by active individuals belonging to the working population and
379 that super-spreader events might be at the origin of such large clusters. Surprisingly, when the
380 level of viral load is analyzed across age classes, no relationship was documented (Supp. Mat.
381 7), meaning that useful information emerges within clusters. Indeed, the characterization of the
382 clusters provides a deeper analysis of the mechanisms behind the progression of an epidemic
383 and the geographic analysis of clusters of cases might constitute a type of investigations to
384 favor in the future.

385

386 ***Non-significant clusters also convey information on the progression of the epidemic***

387 Significant and non-significant clusters both show the same epidemiological trajectories.
388 Indeed, they display similar patterns in terms of increase and decrease in size with a difference
389 in amplitude only. This suggests that the occurrence of clusters, even non-significant, is a good

390 estimator of the epidemic situation. However, significant and non-significant clusters differ in
391 terms of number of cases and measured viral load, but not duration. This suggests (i) that non-
392 significant clusters might correspond to transmission events unrelated to subjects with very
393 high viral load, (ii) that they translate a lower impact on the population in terms of viral spread,
394 and (iii) that they express a transition towards or from a significant spatio-temporal
395 configuration.

396

397 **Limitations**

398 ***Tested population is not homogeneous through time***

399 During the course of the studied epidemic wave, recommendations for testing requested by
400 the authorities have regularly changed. Initially only symptomatic patients at risk and health
401 workers were tested. Then, since mid-March, a wider portion of the population was
402 progressively tested, although younger subjects still were reluctant to get tested. This might
403 have generated heterogeneity in our longitudinal investigation. Moreover, tests could have
404 been done at different stages of the infection (early, late, etc.) and not be representative of the
405 right window of infection. Days of the week might also generate differences in number of
406 positive tests as the number of tests was often reduced over the weekend, for example some
407 persons preferred to be tested only on the next Monday, to avoid quarantine over the week-
408 end.

409

410 ***Positive cases might be missing***

411 Our estimate of the number of positive cases might not be fully representative of the epidemics,
412 particularly at the beginning of the event when only symptomatic patients at risk and health
413 workers were tested. We might assume that their close relatives might also have been infected
414 but were not tested at that time due to reagents shortage. This bias might however have a
415 limited impact on the assessment of cluster size, since any person in contact with the positive
416 cases (documented by the contact tracing team) was tested. Besides, a source of
417 underestimation may be the false negative RT-PCRs, due to imperfect nasopharyngeal

418 sampling. They might however not be of importance as the clinical sensitivity of RT-PCR
419 performed on nasopharyngeal samples in our laboratory is very good, i.e. about 96 to 98%
420 (46, 47). Similarly, the rate of false positives in the same laboratory is estimated to be lower
421 than 1/10'000 tests, thank to full automation and bar-coding that have been settled to prevent
422 human errors and samples/tubes inversion (34). Finally, missing cases might be due to the
423 fact that that some people living or working at the border of the state might have been tested
424 in another state, but again this seems to have a limited impact since more than 80% of all
425 samples tested for SARS-CoV-2 in 2020 have been taken from subjects living in the canton of
426 Vaud.

427

428 **Added value of the methods used**

429 ***Geographic clusters to characterize the epidemics: a key tool for intervention***

430 Beyond the fact that a formal definition of what a cluster is in a geographical context is lacking,
431 the statistical approaches used make implicit assumptions that – through different parameters
432 – have a direct influence on cluster detection and on how to interpret them. We used two
433 complementary approaches that highlight different key aspects of disease clustering. Space-
434 time scan statistics detect the geographical location of case clusters, assess their significance,
435 and characterize their relative risk and duration. This prospective approach is particularly
436 appropriate for the establishment of a daily surveillance system, since it identifies 'alive'
437 clusters only, i.e. having an excess of relative risk on the day of analysis (48). Unlike other
438 detection methods, this approach search for clusters without imposing the specification of their
439 size and allow for analysis of area with heterogeneous population densities. Indeed it identifies
440 a cluster if risk of disease within a space-time cylinder (radius = space, and height = time) is
441 higher than outside. This type of information is key for public health authorities to target
442 neighborhoods and calibrate protective or preventive measures to be deployed.

443 As for the MST-DBSCAN algorithm, it characterizes the diffusion dynamic of the transmission
444 clusters. Here the input parameters require a precise definition of the incubation period, the
445 cluster transmission areas, and a minimum number of spatio-temporal neighbors required to

446 form a cluster (33). The output is a cluster typology according to their behaviors, which is of
447 interest to design sets of appropriate measures to control them. The space-time (SaTScan)
448 and the diffusion-type (MST-DBSCAN) analyses thus provide complementary results in terms
449 of clusters emergence, duration, and demographic characteristics. The two approaches used
450 in conjunction, allow thus for detailed monitoring of the disease's epidemic trajectory and
451 populations at risk and offer adequate tools for governments to both prioritize interventions on
452 excess-risk locations and develop adapted strategies to control cluster diffusion types.

453

454 ***Maps reflect the chronology of the epidemic***

455 The results displayed on static and animated maps well reflect the chronology of the sanitary
456 situation during the first wave of the epidemics. For instance, the major clusters in the Joux
457 valley area can be clearly observed on different maps (Figures 6B, 6B', 6C, 6C'). Noteworthy,
458 these large clusters originate from a super-spreader event that took place end of February in
459 a religious ceremony in Mulhouse, France. Many swiss residents participated in this ceremony
460 and additional related clusters were also observed during the same period north of the
461 Lausanne urban area, and along the Jura mountains (e.g. Morges and Nyon). Conversely,
462 Lausanne was early hit by clusters likely due to a first transmission event that occurred in
463 Northern Italy.

464 Interestingly, the initial phase observed in the state of Vaud differ from what happened in
465 Geneva, where the first clusters emerged in deprived neighborhoods eight days (March 5) after
466 the first positive case (February 26) was detected (22, 8). In Vaud instead, the initial cluster
467 was directly detected the day of the first cases (March 4), with 9 positives and in a wealthy
468 neighborhood.

469

470 ***Positive impact of soft lockdown***

471 The soft lockdown was directly associated with a rapid reduction in the number of positive
472 cases and this despite the increased rate of testing. The reduction takes place massively and
473 in two clear phases in the main urban areas (see Figure 7B), while it happens as a succession

474 of clusters increase and decrease in smaller urban centers and less dense areas (Figure 7D).
475 However, due to the time lag between the identification of positive individuals and the
476 constitution of clusters, the cluster burden occurred directly after the implementation of the soft
477 lockdown. Similarly, the largest clusters, the longest duration and the clusters with individuals
478 showing large viral loads were observed just after the same time-lag. This time lag seems
479 shorter in urban areas as compared to rural areas, likely reflecting the faster spread of the
480 virus in large town such as Lausanne, Morges, Nyon, Yverdon and on the Vaud Riviera. This
481 faster spread is likely due to differences in social and cultural organization between rural and
482 urban areas, to less available room per person in housings, with higher risk of subsequent
483 infection in family with lower socio-economic situations.

484 Our results highlight the efficacy of the lockdown strategy, even soft, to control the epidemic
485 and to decrease the number of positive cases. It also demonstrates the importance of acting
486 when the number of positive cases increases and not waiting for the settlement of clusters.
487 Besides, our results show that the relative risk stayed very low within all the lockdown period.
488 Of note, the compliance of Swiss residents during the first soft lockdown is signaled by the
489 absence of any significant cluster from May 3 to June 16. And finally, it has not escaped our
490 notice that it is already possible to observe the beginnings of the second wave from June 22,
491 2020 (Figure 2A), that is to say exactly two weeks after a series of relaxations of the protective
492 measures such as in particular the authorization of public demonstrations up to 300 people,
493 and the opening of nightclubs (June 6, 2020).

494

495 **Conclusion**

496 Our results highlight that cluster size goes along with the presence of individuals with high viral
497 load, the latter being more commonly found in clusters harboring the youngest group age. This
498 work also stresses the fact that cluster size and cluster duration are largely dependent on the
499 viral load of a few number of individuals within a given cluster, underlying the impact of viral
500 load on contagiousness.

501 Altogether, we provide robust data suggesting that transmission may occur even when all
502 possible source cases in a cluster present a viral load lower than 100'000 copies/ml. Such low
503 viral load cases remain undetected by antigen testing, hence underlying the importance of RT-
504 PCRs assay in case finding and tracing strategies. This in-depth analysis suggests that even
505 older at-risk individuals that try to avoid infection may get infected by SARS-CoV-2 and hence
506 by one of the other cluster members, even when all cluster members exhibit a low viral load,
507 i.e. below 100'000 copies/ml.

508 Finally, such a spatio-temporal characterization of clusters demonstrates the huge effect of the
509 soft lockdown that took place in Switzerland from March 16 to April 27, 2020. Those important
510 results have been documented thanks to the contribution of the geospatial analysis of clusters.

511

512 **Acknowledgments**

513 Anaïs Ladoy is funded by the Direction Générale de la Santé of the canton of Vaud (DGS-
514 Vaud) in the context of the GEOSAN project (Grant Agreement C/20-21 /037).

515

516 **Ethical statement**

517 This study was approved by the Commission cantonale d'éthique de la recherche sur l'être
518 humain (CER-VD), Switzerland. Authorization no. 2020-01302 (20.7.2020).

519

520 **Data availability statement**

521 The dataset analyzed during the current study is available from the corresponding author
522 upon reasonable request. The dataset could not be made publicly available due to the
523 sensitivity of individual georeferenced SARS-CoV-2 testing data. Requests to access the
524 data should be directed to Prof. Gilbert Greub (gilbert.greub@chuv.ch).

525

526

527

528

529 **Author contributions**

530 AL performed the data analyses and drafted the manuscript. AL, SV, SJ and GG conceived
531 the study and completed the manuscript. OO, PNC, and IG participated in the design of the
532 study and helped to draft the manuscript. All authors read and approved the final manuscript.

533

534 **Conflict of interest**

535 The authors declare to have no conflict of interest.

536

537 **References**

- 538 1. World Health Organization. 2021. Weekly epidemiological update - 12 January 2021.
539 Wkly Epidemiol Update - 12 January 2021.
- 540 2. World Health Organization. 2021. WHO Coronavirus Disease (COVID-19) Dashboard.
- 541 3. Nicola M, Sohrabi C, Mathew G, Kerwan A, Al-Jabir A, Griffin M, Agha M, Agha R.
542 2020. Health policy and leadership models during the COVID-19 pandemic: A review.
543 Int J Surg 81:122–129.
- 544 4. Chu IY-H, Alam P, Larson HJ, Lin L. 2020. Social consequences of mass quarantine
545 during epidemics: a systematic review with implications for the COVID-19 response. J
546 Travel Med 27.
- 547 5. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, Sanderson C,
548 McKee M, Troeger C, Ong KL, Checchi F, Perel P, Joseph S, Gibbs HP, Banerjee A,
549 Eggo RM, Nightingale ES, O'Reilly K, Jombart T, Edmunds WJ, Rosello A, Sun FY,
550 Atkins KE, Bosse NI, Clifford S, Russell TW, Deol AK, Liu Y, Procter SR, Leclerc QJ,
551 Medley G, Knight G, Munday JD, Kucharski AJ, Pearson CAB, Klepac P, Prem K,
552 Houben RMGJ, Endo A, Flasche S, Davies NG, Diamond C, van Zandvoort K, Funk S,
553 Auzenbergs M, Rees EM, Tully DC, Emery JC, Quilty BJ, Abbott S, Villabona-Arenas
554 CJ, Hué S, Hellewell J, Gimma A, Jarvis CI. 2020. Global, regional, and national
555 estimates of the population at increased risk of severe COVID-19 due to underlying
556 health conditions in 2020: a modelling study. Lancet Glob Health 8:e1003–e1017.

- 557 6. Ruktanonchai NW, Floyd JR, Lai S, Ruktanonchai CW, Sadilek A, Rente-Lourenco P,
558 Ben X, Carioli A, Gwinn J, Steele JE, Prosper O, Schneider A, Oplinger A, Eastham P,
559 Tatem AJ. 2020. Assessing the impact of coordinated COVID-19 exit strategies across
560 Europe. *Science* 369:1465–1470.
- 561 7. Faber M, Ghisletta A, Schmidheiny K. 2020. A lockdown index to assess the economic
562 impact of the coronavirus. *Swiss J Econ Stat* 156:11.
- 563 8. De Ridder D, Sandoval J, Vuilleumier N, Stringhini S, Spechbach H, Joost S, Kaiser L,
564 Guessous I. 2020. Geospatial digital monitoring of COVID-19 cases at high
565 spatiotemporal resolution. *Lancet Digit Health* 2:e393–e394.
- 566 9. Hohl A, Delmelle EM, Desjardins MR, Lan Y. 2020. Daily surveillance of COVID-19
567 using the prospective space-time scan statistic in the United States. *Spat Spatio-
568 Temporal Epidemiol* 34:100354.
- 569 10. Franch-Pardo I, Napoletano BM, Rosete-Verges F, Billa L. 2020. Spatial analysis and
570 GIS in the study of COVID-19. A review. *Sci Total Environ* 739:140033.
- 571 11. Keesara S, Jonas A, Schulman K. 2020. Covid-19 and Health Care’s Digital
572 Revolution. *N Engl J Med* 382:e82.
- 573 12. Cromley EK. 2019. Using GIS to Address Epidemiologic Research Questions. *Curr
574 Epidemiol Rep* 6:162–173.
- 575 13. Public Health England. 2020. COVID-19: epidemiological definitions of outbreaks and
576 clusters in particular settings. GOVUK.
- 577 14. Lessler J, Azman AS, McKay HS, Moore SM. 2017. What is a Hotspot Anyway? *Am J
578 Trop Med Hyg* 96:1270–1273.
- 579 15. UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance, Joint United
580 Nations Programme on HIV/AIDS, World Health Organization. 2013. Guidelines for
581 second generation HIV surveillance: an update : know your epidemic.
- 582 16. Bejon P, Williams TN, Liljander A, Noor AM, Wambua J, Ogada E, Olotu A, Osier FHA,
583 Hay SI, Färnert A, Marsh K. 2010. Stable and Unstable Malaria Hotspots in
584 Longitudinal Cohort Studies in Kenya. *PLOS Med* 7:e1000304.

- 585 17. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, Ghani A,
586 Drakeley C, Gosling R. 2012. Hitting Hotspots: Spatial Targeting of Malaria for Control
587 and Elimination. *PLOS Med* 9:e1001165.
- 588 18. Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. 1996. Spatial Point Pattern Analysis
589 and Its Application in Geographical Epidemiology. *Trans Inst Br Geogr* 21:256–274.
- 590 19. Jacquez GM, Greiling DA. 2003. Local clustering in breast, lung and colorectal cancer
591 in Long Island, New York. *Int J Health Geogr* 2:3.
- 592 20. Zhang X, Rao H, Wu Y, Huang Y, Dai H. 2020. Comparison of spatiotemporal
593 characteristics of the COVID-19 and SARS outbreaks in mainland China. *BMC Infect*
594 *Dis* 20:805.
- 595 21. Desjardins MR, Hohl A, Delmelle EM. 2020. Rapid surveillance of COVID-19 in the
596 United States using a prospective space-time scan statistic: Detecting and evaluating
597 emerging clusters. *Appl Geogr* 118:102202.
- 598 22. De Ridder D, Sandoval J, Vuilleumier N, Azman AS, Stringhini S, Kaiser L, Joost S,
599 Guessous I. 2020. Socioeconomically disadvantaged neighborhoods face increased
600 persistence of SARS-CoV-2 clusters. *Front Public Health* 8.
- 601 23. Kamel Boulos MN, Geraghty EM. 2020. Geographical tracking and mapping of
602 coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2
603 (SARS-CoV-2) epidemic and associated events around the world: how 21st century
604 GIS technologies are supporting the global fight against outbreaks and epidemics. *Int J*
605 *Health Geogr* 19:8.
- 606 24. Danis K, Epaulard O, Bénet T, Gaymard A, Campoy S, Botelho-Nevers E,
607 Bouscambert-Duchamp M, Spaccaferri G, Ader F, Mailles A, Boudalaa Z, Tolsma V,
608 Berra J, Vaux S, Forestier E, Landelle C, Fougere E, Thabuis A, Berthelot P, Veil R,
609 Levy-Bruhl D, Chidiac C, Lina B, Coignard B, Saura C, Investigation Team. 2020.
610 Cluster of Coronavirus Disease 2019 (COVID-19) in the French Alps, February 2020.
611 *Clin Infect Dis* 71:825–832.

- 612 25. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005. Superspreading and the
613 effect of individual variation on disease emergence. *Nature* 438:355–359.
- 614 26. Stein RA. 2011. Super-spreaders in infectious diseases. *Int J Infect Dis* 15:e510–e513.
- 615 27. Frieden TR, Lee CT. 2020. Identifying and Interrupting Superspreading Events-
616 Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg*
617 *Infect Dis* 26:1059–1066.
- 618 28. Beldomenico PM. 2020. Do superspreaders generate new superspreaders? A
619 hypothesis to explain the propagation pattern of COVID-19. *Int J Infect Dis* 96:461–
620 463.
- 621 29. Jacot D, Greub G, Jatton K, Opota O. 2020. Viral load of SARS-CoV-2 across patients
622 and compared to other respiratory viruses. *Microbes Infect* 22:617–621.
- 623 30. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. 2020. The reproductive number of COVID-
624 19 is higher compared to SARS coronavirus. *J Travel Med* 27.
- 625 31. Kulldorff M. 1997. A spatial scan statistic. *Commun Stat - Theory Methods* 26:1481–
626 1496.
- 627 32. Moraga P, Montes F. 2011. Detection of spatial disease clusters with LISA functions.
628 *Stat Med* 30:1057–1071.
- 629 33. Kuo F-Y, Wen T-H, Sabel CE. 2018. Characterizing Diffusion Dynamics of Disease
630 Clustering: A Modified Space–Time DBSCAN (MST-DBSCAN) Algorithm. *Ann Am*
631 *Assoc Geogr* 108:1168–1186.
- 632 34. Greub G, Sahli R, Brouillet R, Jatton K. 2016. Ten years of R&D and full automation in
633 molecular diagnosis. *Future Microbiol* 11:403–425.
- 634 35. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink
635 S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, Veer B van der, Brink S van
636 den, Wijsman L, Goderski G, Romette J-L, Ellis J, Zambon M, Peiris M, Goossens H,
637 Reusken C, Koopmans MP, Drosten C. 2020. Detection of 2019 novel coronavirus
638 (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25:2000045.

- 639 36. Pillonel T, Scherz V, Jatton K, Greub G, Bertelli C. 2020. Letter to the editor: SARS-
640 CoV-2 detection by real-time RT-PCR. Euro Surveill Bull Eur Sur Mal Transm Eur
641 Commun Dis Bull 25.
- 642 37. Opota O, Brouillet R, Greub G, Jatton K. 2020. Comparison of SARS-CoV-2 RT-PCR
643 on a high-throughput molecular diagnostic platform and the cobas SARS-CoV-2 test for
644 the diagnostic of COVID-19 on various clinical samples. Pathog Dis 78.
- 645 38. Moraz M, Jacot D, Papadimitriou-Olivgeris M, Senn L, Greub G, Jatton K, Opota O.
646 2020. Universal admission screening strategy for COVID-19 highlighted the clinical
647 importance of reporting SARS-CoV-2 viral loads. New Microbes New Infect 38:100820.
- 648 39. SFSO. 2019. STATPOP - Population and Households Statistics. Fed Stat Off.
- 649 40. SFSO. 2020. Population and Household Statistics (STATPOP), spatial data 2019. Fed
650 Stat Off.
- 651 41. Besag J, Diggle PJ. 1977. Simple Monte Carlo Tests for Spatial Pattern. J R Stat Soc
652 Ser C Appl Stat 26:327–333.
- 653 42. Ester M, Kriegel H-P, Xu X. 1996. A Density-Based Algorithm for Discovering Clusters
654 in Large Spatial Databases with Noise, p. 6. In Proceedings of the Second International
655 Conference on Knowledge Discovery and Data Mining Evangelos Simoudis, Jiawei
656 Han, and Usama Fayyad. AAAI Press, Menlo Park, California.
- 657 43. MicroGIS S. 2019. LC Swiss Localities - Swiss postcode database.
- 658 44. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of
659 communities in large networks. J Stat Mech Theory Exp 2008:P10008.
- 660 45. Caruana G, Croxatto A, Kampouri E, Kritikos A, Opota O, Foerster M, Brouillet R, Senn
661 L, Lienhard R, Egli A, Pantaleo G, Carron P-N, Greub G. 2021. ImplemeNting SARS-
662 CoV-2 Rapid antigen testing in the Emergency wArd of a Swiss univErsity hospital: the
663 INCREASE study. medRxiv 2021.02.10.21250915.
- 664 46. Mueller L, Scherz V, Greub G, Jatton K, Opota O. 2020. Computer-aided medical
665 microbiology monitoring tool: a strategy to adapt to the SARS-CoV-2 epidemic and that
666 highlights RT-PCR consistency. medRxiv 2020.07.27.20162123.

- 667 47. Schwob JM, Miauton A, Petrovic D, Perdrix J, Senn N, Jatton K, Onya O, Maillard A,
668 Minghelli G, Cornuz J, Greub G, Genton B, D'Acremont V. 2020. Antigen rapid tests,
669 nasopharyngeal PCR and saliva PCR to detect SARS-CoV-2: a prospective
670 comparative clinical trial. medRxiv 2020.11.23.20237057.
- 671 48. Kulldorff M. 2001. Prospective time periodic geographical disease surveillance using a
672 scan statistic. J R Stat Soc Ser A Stat Soc 164:61–72.

673

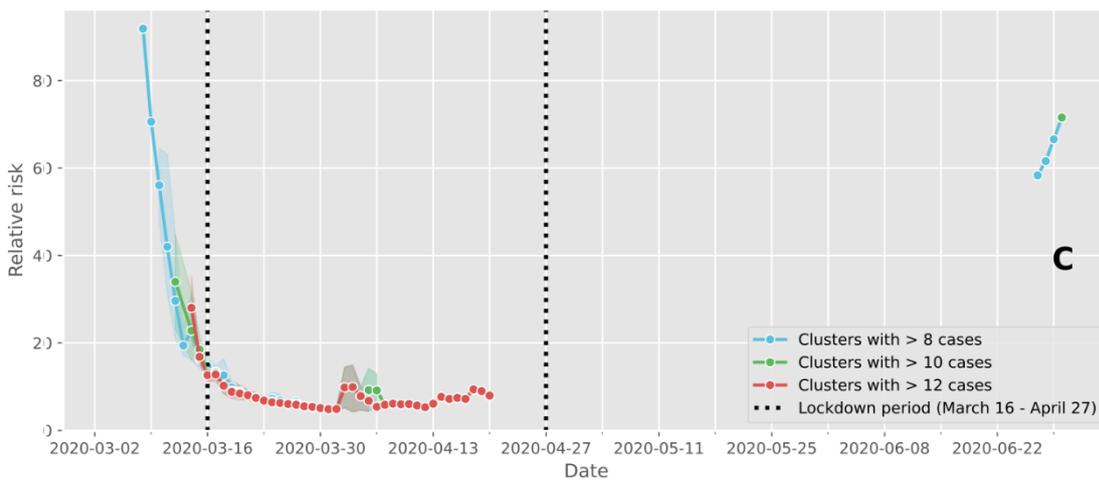
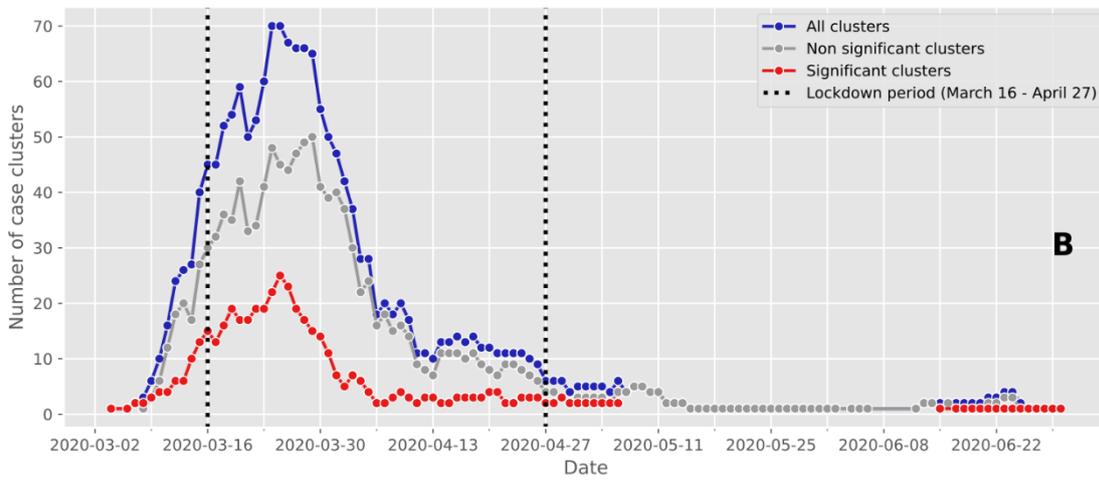
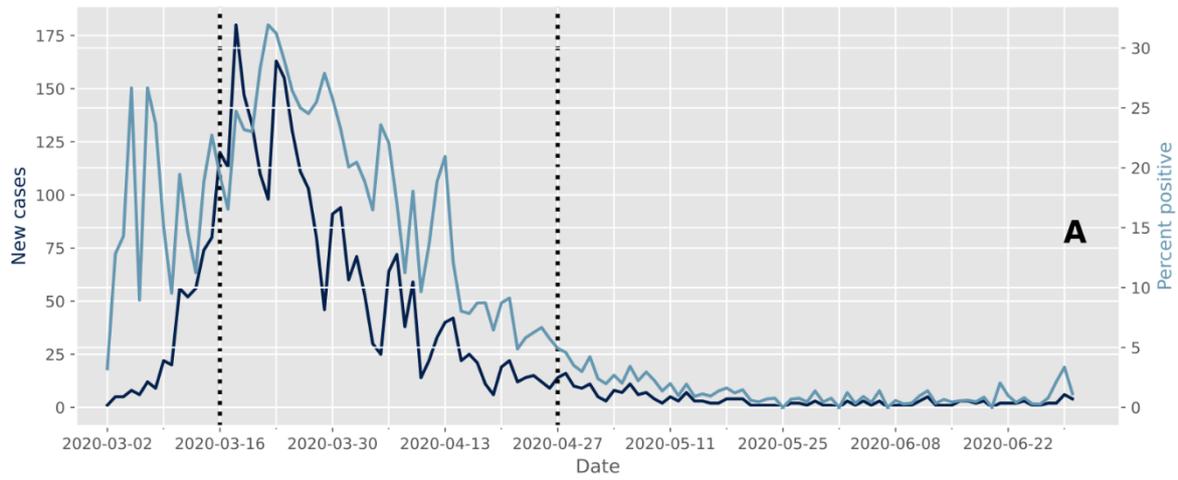
674

675 **Figures**

676

677 **Figure 1.** Evolution of cases and clusters through time.

678 On the figures, the vertical dashed lines delimit the Swiss lockdown period (March 16 – April
679 27). (A) Epidemic trajectory of positive tests. The daily new confirmed cases are represented
680 in dark blue whereas the percent of positive tests are shown in light blue. The peak of daily
681 new cases occurred on March 18, while the highest proportion of positive tests was recorded
682 on March 22. (B) Number of case clusters over time. The total number of clusters detected
683 daily by space-time scan statistics is shown in dark blue, while the red and gray lines represent
684 the proportion of significant clusters ($p \leq 0.05$) and non significant clusters ($p > 0.05$) respectively.
685 (C) Average relative risk of significant space-time clusters ($p \leq 0.05$) over time according to the
686 within-cluster cases. As the expected number of cases was very low in some sparsely
687 populated rural areas, resulting in extremely high relative risk values, we calculated the relative
688 risk by considering only clusters with more than 8 (blue line), 10 (green line) and 12 cases (red
689 line).

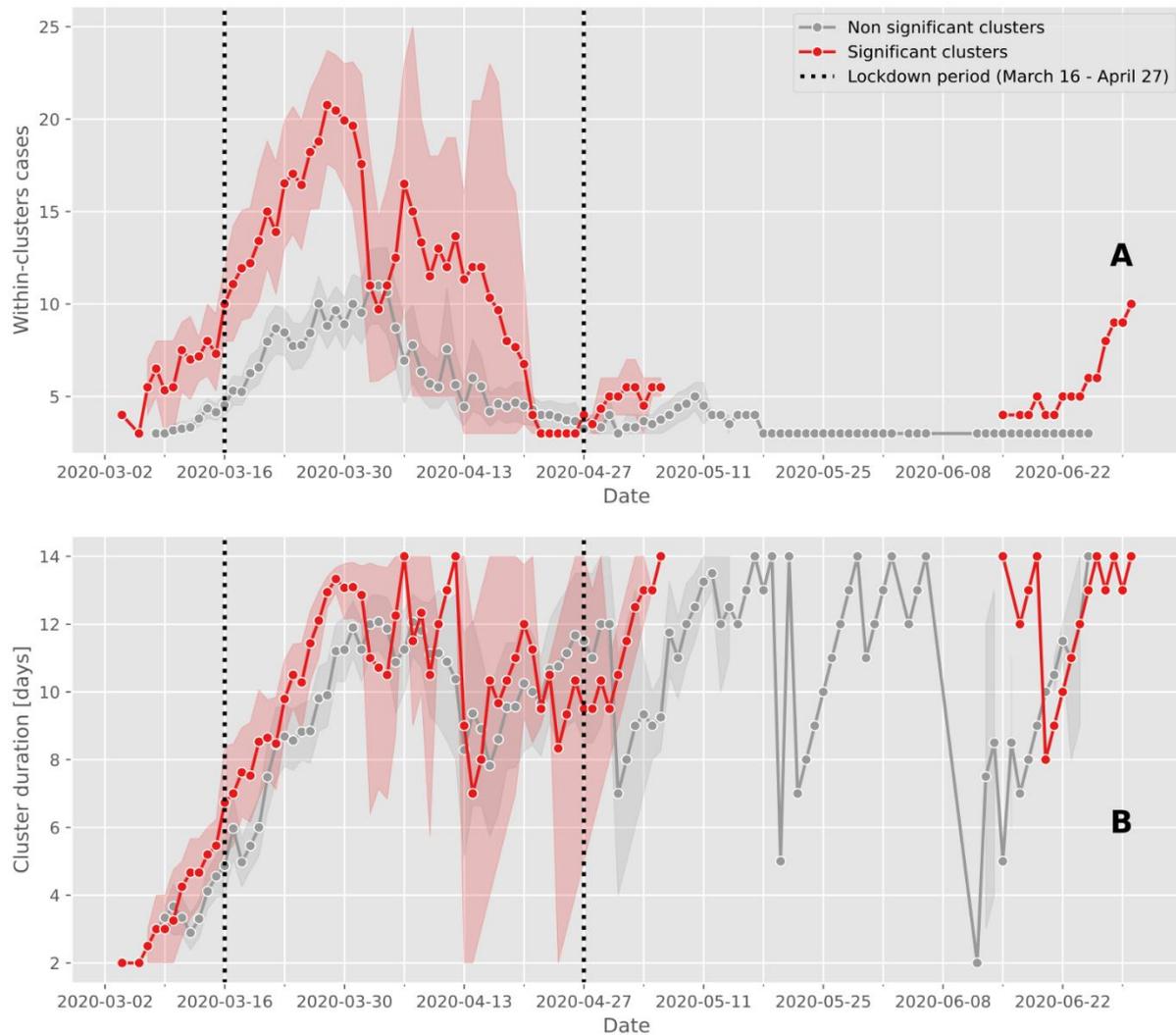


690

691

692 **Figure 2.** Case cluster characteristics over time.

693 The average and confidence intervals of the number of cases within clusters (A) and the
694 duration of clusters (B) are calculated for significant (red line) and non significant (grey
695 line). Between May 6 and June 15, the prospective space-time scan statistic detected no
696 significant clusters. In the figures, the vertical dashed lines delimit the Swiss lockdown period
697 (March 16 – April 27).



698

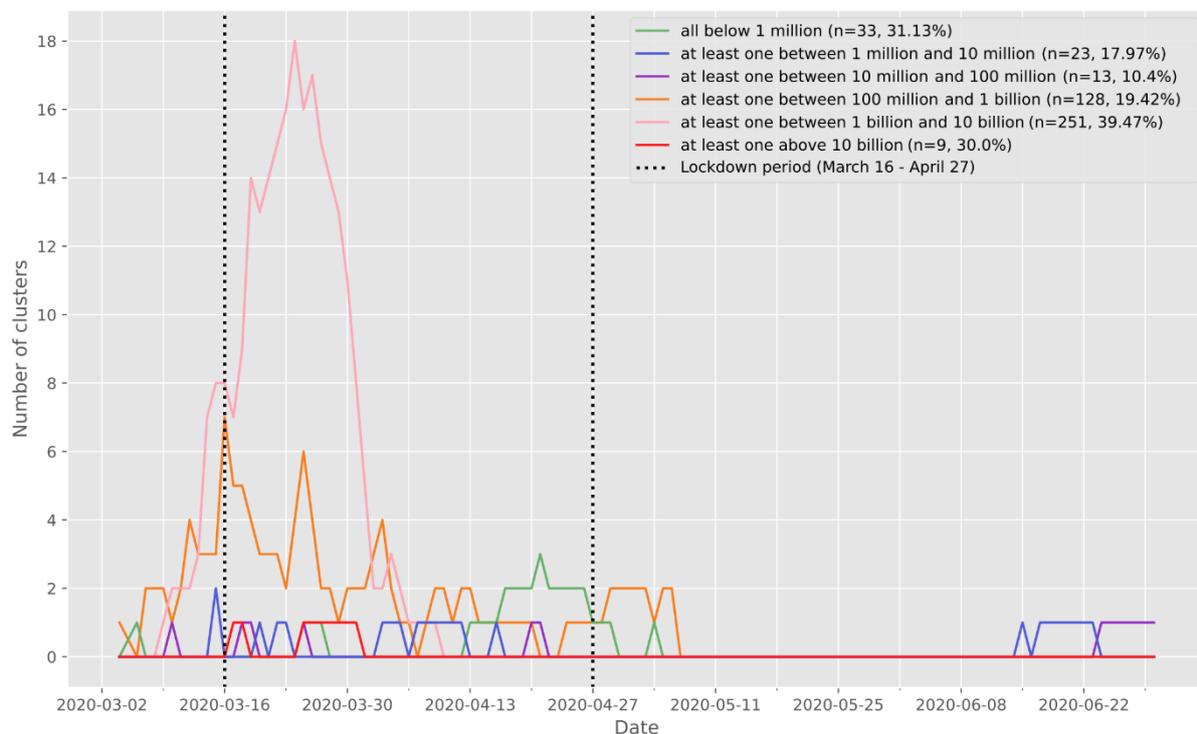
699 **Table 1.** Classification of the space-time clusters according to the viral load of the cases
700 involved.
701 Within-cluster cases were identified by matching both geographically and temporally positive
702 test subjects, geocoded at the residential address, with space-time clusters. For example, a
703 cluster was classified as "all below 1 million" if all individuals tested positive within the cluster
704 during its active period had a viral load below 1 million copies/ml. Noteworthy, in 20 significant
705 clusters the viral load of three first cases were all below 100'000 copies/ml.
706 For each cluster category, the total number of case clusters detected by prospective space-
707 time scan statistics over the entire study period (March 2-June 20) and the proportion of
708 significant clusters ($p \leq 0.05$) are reported.

Case cluster viral load category	Total number of clusters	Number of significant clusters
All below 1 million	106	33 (31.13%)
At least one between 1 million and 10 million	128	23 (17.97%)
At least one between 10 million and 100 million	125	13 (10.40%)
At least one between 100 million and 1 billion	659	128 (19.42%)
At least one between 1 billion and 10 billion	636	251 (39.47%)
At least one above 10 billion	30	9 (30.00%)
<i>Total</i>	1684	457

709

710

711 **Figure 3.** Number of significant ($p \leq 0.05$) case clusters over time characterized according to
712 the viral load of the cases documented in each cluster.
713 The vertical dashed lines delimit the Swiss lockdown period (March 16 – April 27). For each
714 cluster, we extracted the positive test individuals intersecting the cluster both geographically
715 and temporally, and we characterized the clusters according to the viral load of the individuals
716 composing it. The clusters represented in green are therefore composed solely of individuals
717 with a viral load of less than 1 million copies/ml. The clusters shown in blue are made up of at
718 least one individual with a viral load between 1 million and 10 million copies/ml, and so on.

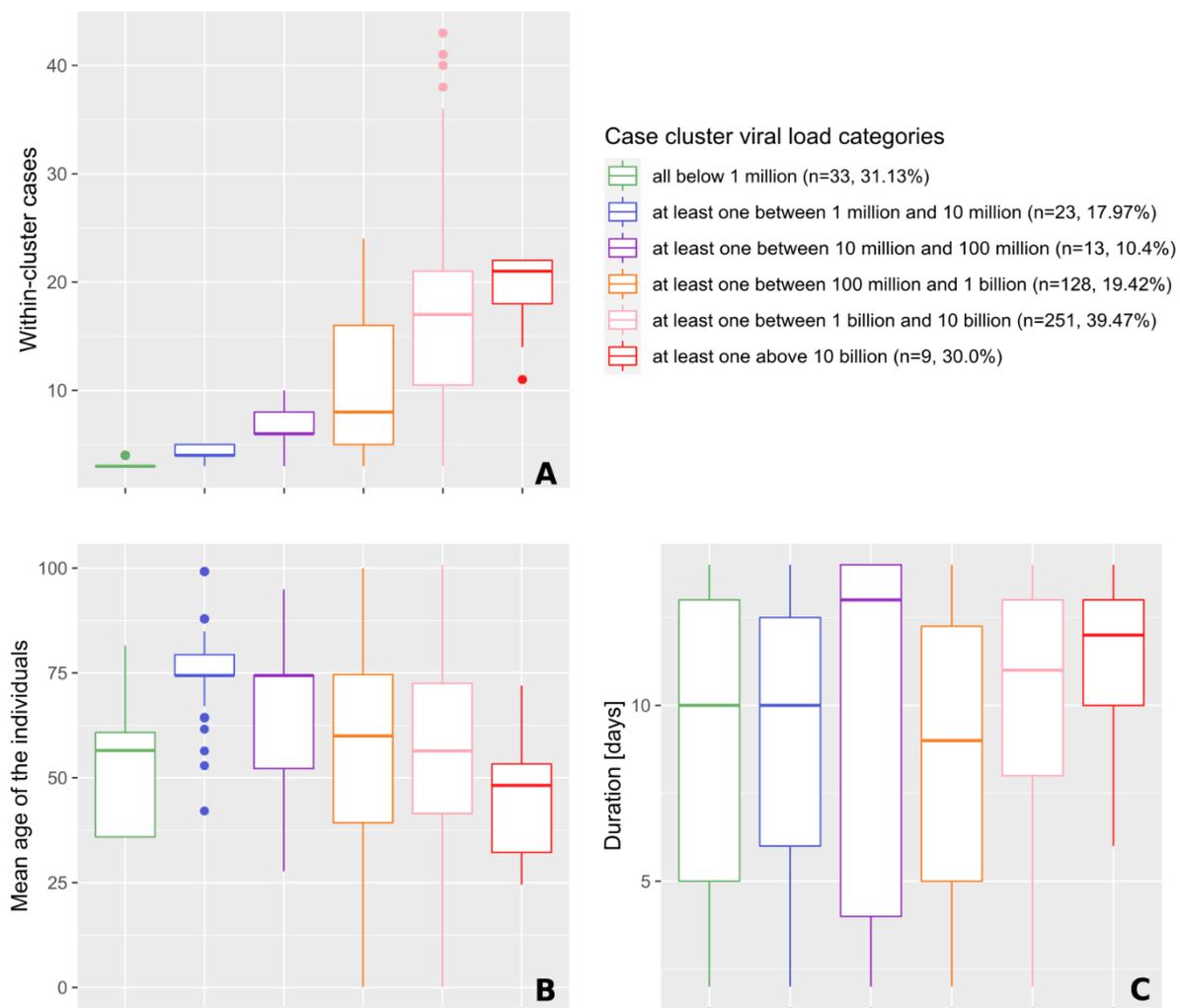


719

720

721

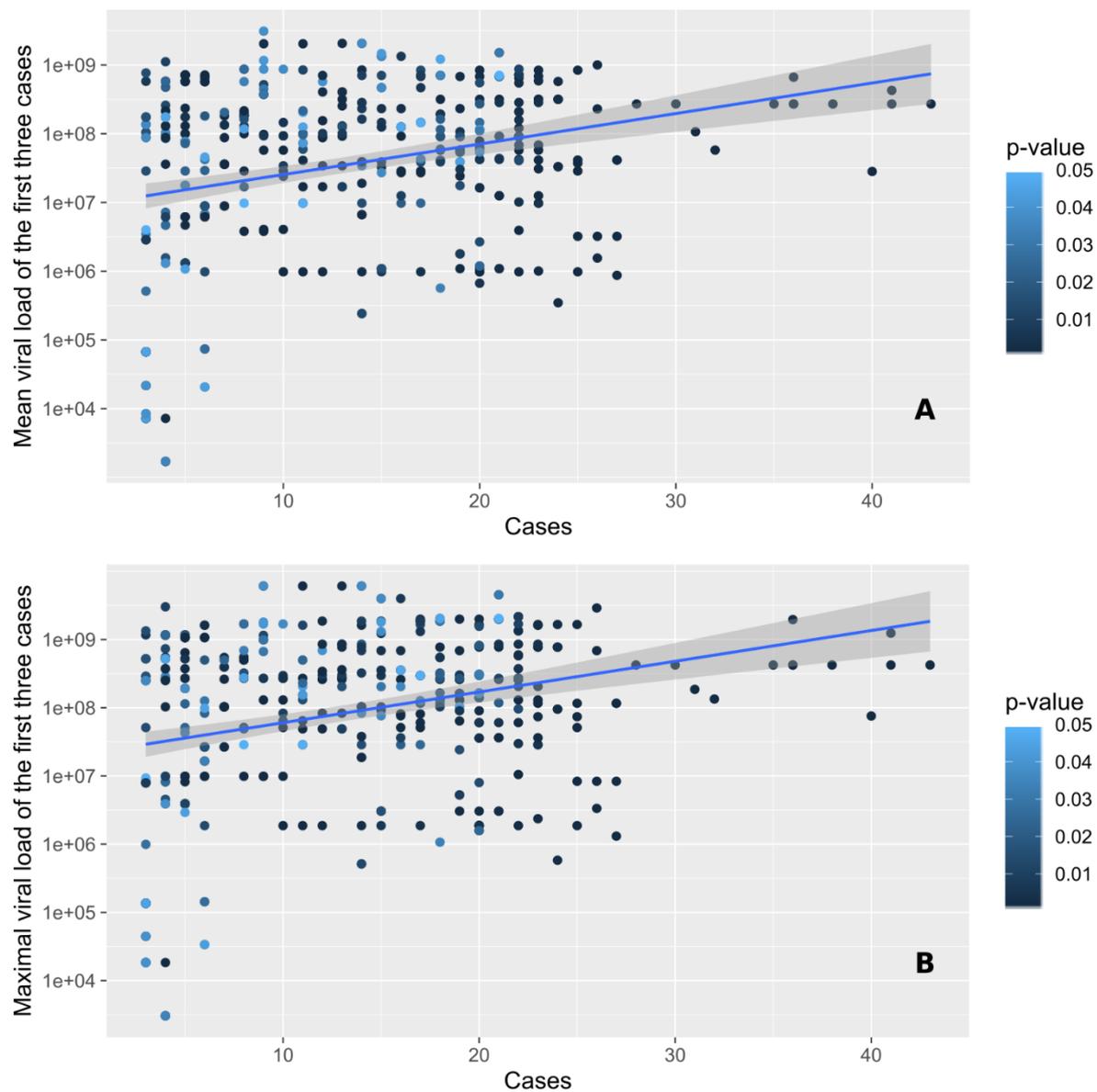
722 **Figure 4.** Characteristics (size, age and duration) of significant space-time clusters ($p \leq 0.05$)
723 over the study period, categorized according to the viral load of the cases involved.
724 The classification procedure is explained into more details in the legend of Table 1.
725 Characteristics include the number of cases observed in clusters (A), the mean age of the
726 positive tests individuals forming clusters (B), and the duration of clusters (C). Of note, the
727 median number of cases was significantly higher (17 individuals) among clusters with subjects
728 showing extremely high viral load > 1 billion copies/ml) as compared to clusters with individuals
729 exhibiting a viral load < 1 million copies/ml (xx individuals).



730

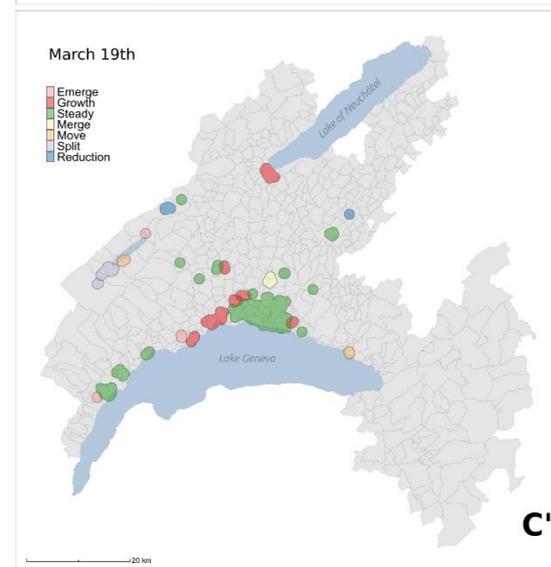
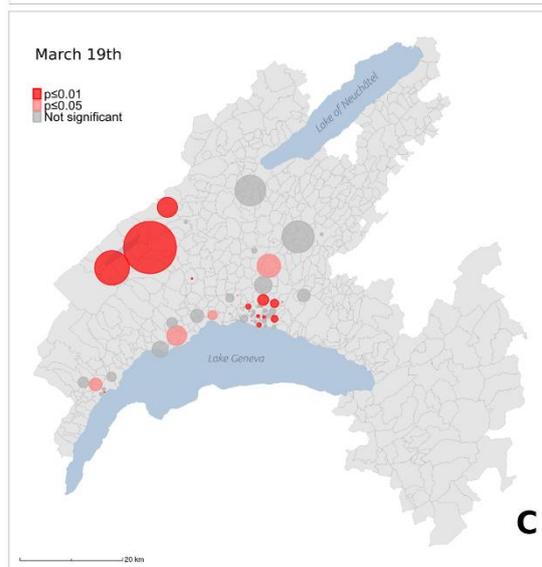
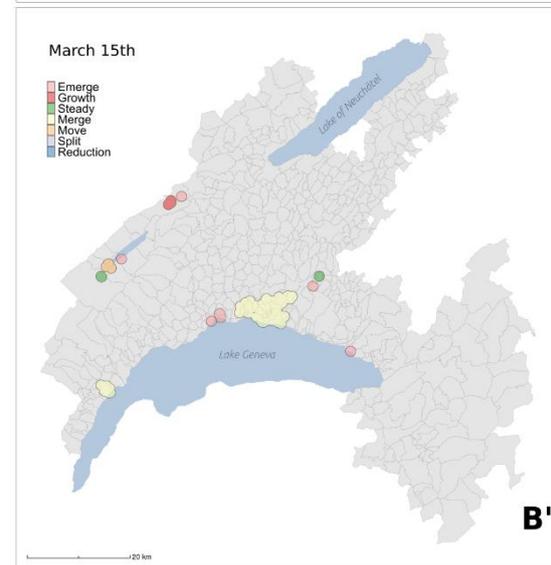
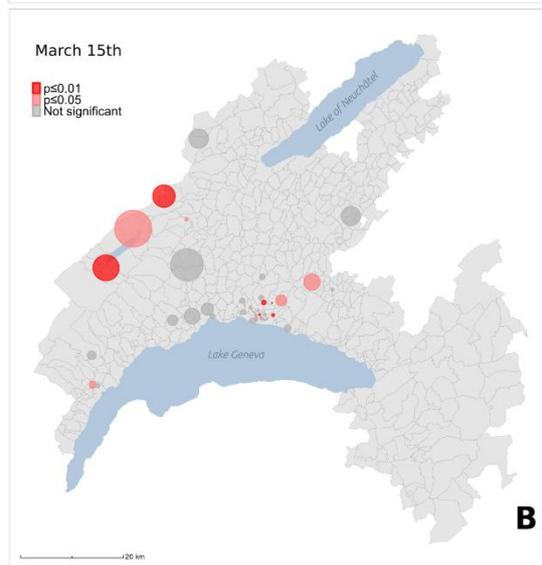
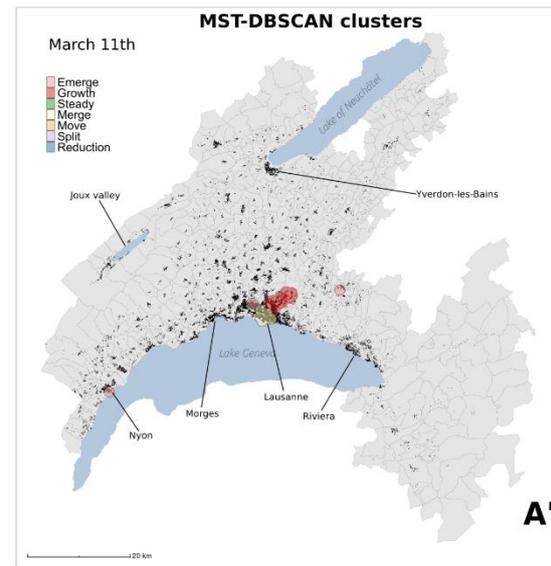
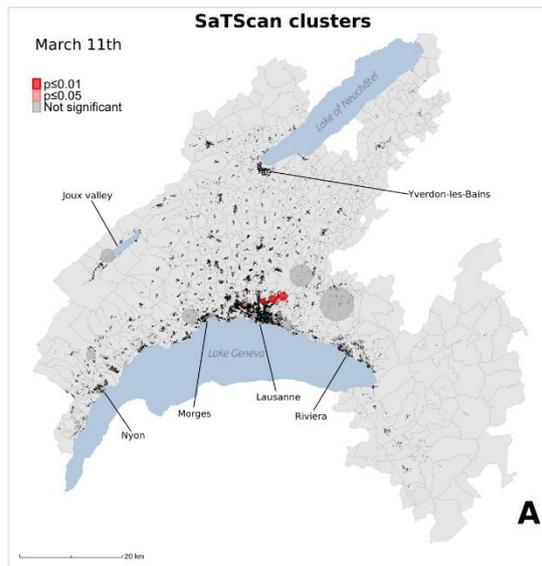
731

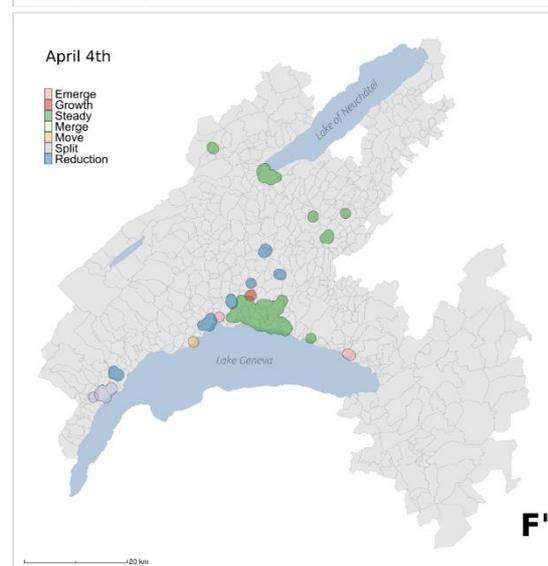
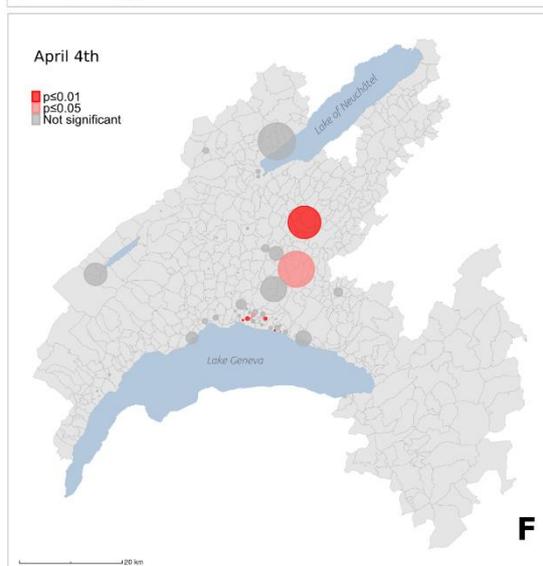
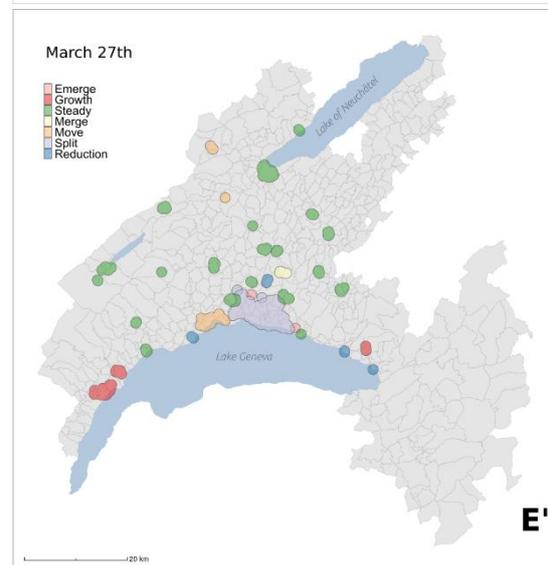
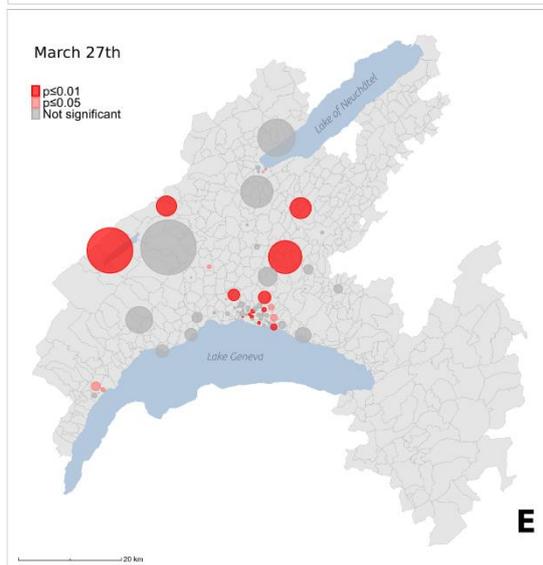
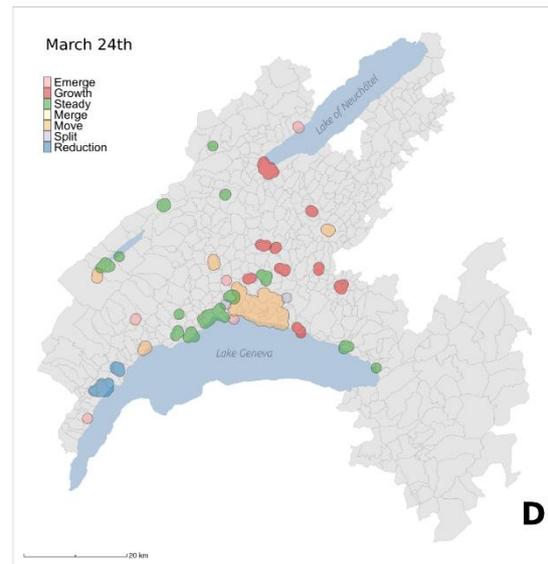
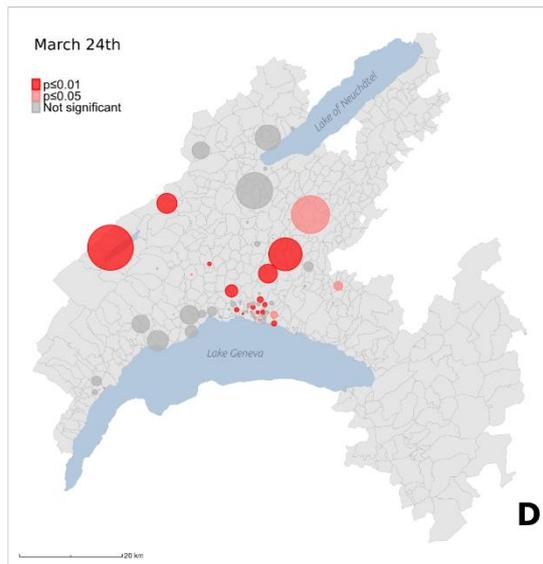
732 **Figure 5.** Number of cases observed within significant space-time clusters ($p \leq 0.05$) in function
733 of the mean (A) and maximal (B) viral load of the first three cases involved. Points are colored
734 according to the significance level of the cluster, which was assessed through 999 Monte Carlo
735 random permutations.



736

737 **Figure 6.** Spatial distribution of case clusters (A-F) and diffusion dynamics of transmission
738 clusters (A'-F') for 6 key dates (March 11, March 15, March 19, March 24, March 27, April 4)
739 during the first epidemic wave.
740 Case clusters resulting from the prospective Poisson space-time scan statistics (A-F) are
741 shaded according to their significance level: dark red for statistically significant clusters with
742 $\alpha = 0.01$, light red for statistically significant clusters with $\alpha = 0.05$, and grey for non
743 significant clusters ($p > 0.05$). Transmission clusters resulting from the MST-DBSCAN algorithm
744 (A'-F') are shaded according to their evolution type: emerge (pink), growth (red), steady
745 (green), merge (yellow), move (orange), split (purple), and reduction (blue).
746 Black points in (A) and (A') represent the 33'651 individuals tested during the study period
747 (January 10 - June 30).
748



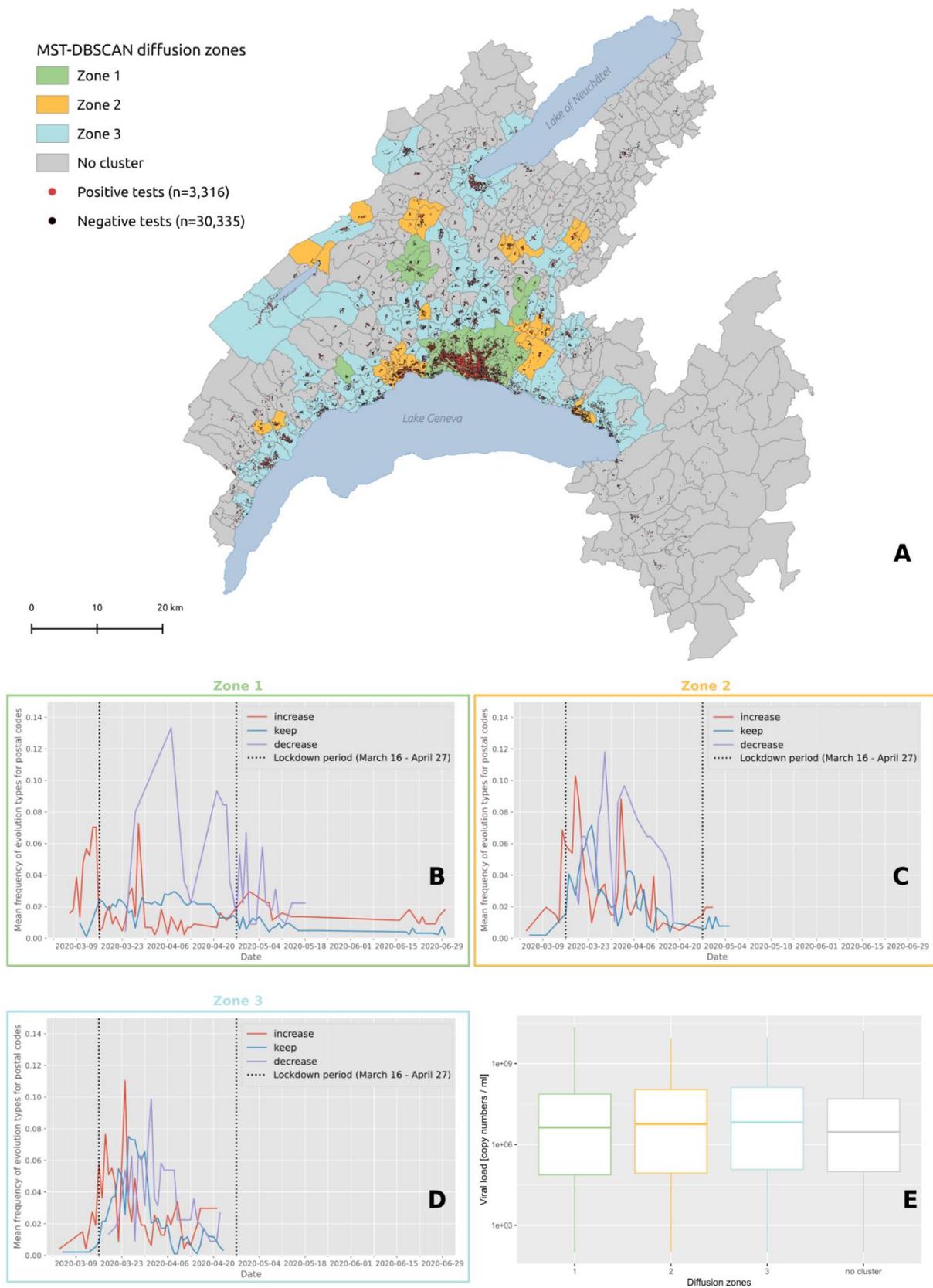


751 **Figure 7.** (A) Diffusion zones identified by the MST-DBSCAN algorithm. Postcode areas with
752 the same color share similar diffusion patterns, and areas without any transmission clusters
753 are represented in grey. The red and black dots on the map indicate the individuals tested
754 positive (n=3'316) and negative (n=30'335) for COVID-19 in the canton of Vaud between
755 January 10 and June 30.

756 Below the map are shown the frequencies of major evolution types over time for zone 1 (B),
757 zone 2 (C) and zone 3 (D). The red line corresponds to the "increase" diffusion type
758 assigned to the transmission clusters whose area becomes larger, the blue line corresponds
759 to the "keep" diffusion type assigned to the transmission clusters whose area remains, and the
760 purple line corresponds to the "decrease" diffusion type assigned to the transmission clusters
761 whose area becomes smaller.

762 The vertical dashed lines delimit the Swiss lockdown period (March 16 – April 27).

763 (E) Distribution of the viral load of test-confirmed cases living in each diffusion zone.



764

765 **Box 1: The first SARS-CoV-2 epidemic wave in the state of Vaud, Switzerland**

766 On March 11, 7 days after the first detection of a positive case (Figure 6A), we observe a phase
767 of rapid growth and merging (see Figure 6A') and a series of significant clusters explodes
768 directly north of Lausanne, the main city of the state. Interestingly three out of the four clusters
769 shown are located in wealthy areas. March 15 (Figures 6B and 6B') is the day before the soft-
770 lockdown. There is a multitude of clusters in the Lausanne area, among which a large fraction
771 is significant (Figure 6B). Figure 6B' shows that these clusters rapidly merged into a single
772 "super cluster" deployed over the urban agglomeration. In the country-side, active clusters
773 emerge and grow north of the lake in the Joux valley located in the Jura mountains where
774 population density is low. Four days later, March 19 (Figures 6C and 6C'), the peak of the first
775 wave is approaching (see Figure 1B). The number of case clusters is high in the Lausanne
776 area (Figure 6C) but clearly stabilizes. Similar behavior is observed towards east, along
777 Geneva lake; in the Riviera area only one moving cluster is observed (Figure 6C'), while new
778 clusters grow in the Morges area. In the Joux valley the activity remains important, and a
779 cluster grows in Yverdon-les-Bains, south of the Lake of Neuchâtel. On March 24 (Figures 6D
780 and 6D'), the peak of the first wave is reached (see Figure 1B). New cases reactivate moving
781 clusters in the center of Lausanne, while towards west the situation stabilizes and even
782 reduces toward Geneva with no more significant cluster in the Nyon area. At the peak, a large
783 significant cluster remains steady in the Jura, and several clusters grow in the remote, rural
784 periphery north of the main urban area (Figures 6D and 6D'). In the north, close to the Lake of
785 Neuchâtel, the clusters are not significant despite growing. March 27 (Figures 6E and 6E') is
786 the start of the important and rapid reduction phase of all clusters (see Figure 1B). The merged
787 clusters of the Lausanne area split and most of those located on the country-side are steady.
788 However, a significant cluster emerges in the west, in the Nyon area. On April 4 (Figures 6F
789 and 6F'), it is the end of the peak. The Joux valley cluster is over after 25 days and clusters in
790 the canton are either not significant anymore, or are steady, split or reduce. There is one
791 exception north of Lausanne with a single growing cluster located in a leisure area and likely
792 related to the presence of a school (with boarding).