




OPEN

## Impact of US vaccination strategy on COVID-19 wave dynamics

Corentin Cot<sup>1,2</sup>, Giacomo Cacciapaglia<sup>1,2</sup>, Anna Sigridur Islind<sup>3</sup>, María Óskarsdóttir<sup>3</sup> & Francesco Sannino<sup>4,5</sup>

We employ the epidemic Renormalization Group (eRG) framework to understand, reproduce and predict the COVID-19 pandemic diffusion across the US. The human mobility across different geographical US divisions is modelled via open source flight data alongside the impact of social distancing for each such division. We analyse the impact of the vaccination strategy on the current pandemic wave dynamics in the US. We observe that the ongoing vaccination campaign will not impact the current pandemic wave and therefore strict social distancing measures must still be enacted. To curb the current and the next waves our results indisputably show that vaccinations alone are not enough and strict social distancing measures are required until sufficient immunity is achieved. Our results are essential for a successful vaccination strategy in the US.

The United States (US), raged by the SARS-CoV-2 virus, are paying an immense toll in terms of the loss of human lives and jobs, with a dreadful impact on society and economy. Understanding and predicting the time evolution of the pandemic plays a key role in defining prevention and control strategies. Short-term forecasts have been obtained, since the early days, via effective methods<sup>1–3</sup>. Furthermore, time-honored mathematical models can be used, like compartmental models<sup>4–8</sup> of the SIR type<sup>9</sup> or complex networks<sup>10–12</sup>. Nevertheless, it remains very hard to understand and forecast the wave pattern of pandemics like COVID-19<sup>13</sup>.

In this work, we employ the *epidemic Renormalization Group* (eRG) framework, recently developed in<sup>14,15</sup>. It can be mapped<sup>15,16</sup> into a time-dependent compartmental model of the SIR type<sup>9</sup>. The eRG framework provides a single first order differential equation, apt to describing the time-evolution of the cumulative number of infected cases in an isolated region<sup>14</sup>. It has been extended in Ref.<sup>15</sup> to include interactions among multiple regions of the world. The main advantage over SIR models is its simplicity, and the fact that it relies on symmetries of the system instead of a detailed description. As a result, no computer simulation is needed in order to understand the time-evolution of the epidemic even at large scales<sup>15</sup>. Recently, the framework has been extended to include the multi-wave pattern<sup>17,18</sup> observed in the COVID-19 and other pandemics<sup>19</sup>.

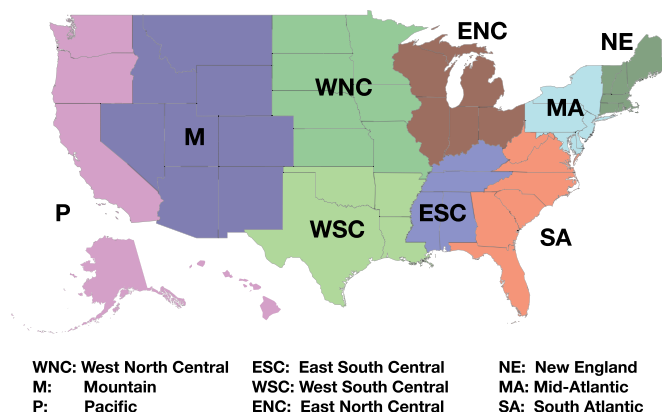
The Renormalization Group approach<sup>20,21</sup> has a long history in physics with impact from particle to condensed matter physics and beyond. Its application to epidemic dynamics is complementary to other approaches<sup>10–12,22–29</sup>. Here we demonstrate that the framework is able to reproduce and predict the pandemic diffusion in the US taking into account the human mobility across different geographical US divisions, as well as the impact of social distancing within each one. To gain an insight and to better monitor the human exchange we make use of open source flight data among the states. We calibrate the model on the first wave pandemic, raging from March to August, 2020. With these insights, we then analyse and understand the current second wave, raging in all the divisions. The eRG framework can also be easily adapted to take into account vaccinations<sup>23</sup>. We propose a new framework and use it to quantify the impact of the vaccination campaign, started on December 14th, on the current and future wave dynamics. Our results are in agreement with previous work based on compartmental models<sup>30</sup>, and confirm that the current campaign will have limited impact on the ongoing wave.

### Methodology

In this section we briefly review our methods that include the open source flight data description, their interplay with the eRG mathematical model framework and, last but not least, the interplay with vaccine deployment and implementation.

<sup>1</sup>Institut de Physique des deux Infinis de Lyon (IP2I), UMR5822, CNRS/IN2P3, 69622 Villeurbanne, France. <sup>2</sup>University of Lyon, Université Claude Bernard Lyon 1, 69001 Lyon, France. <sup>3</sup>Department of Computer Science, Reykjavík University, Menntavegur 1, 102 Reykjavík, Iceland. <sup>4</sup>CP3-Origins & the Danish Institute for Advanced Study, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark. <sup>5</sup>Dipartimento di Fisica E. Pancini, Università di Napoli Federico II & INFN sezione di Napoli, Complesso Universitario di Monte S. Angelo Edificio 6, via Cintia 80126, Napoli, Italy. ✉email: cot@ipnl.in2p3.fr

## Geographical Divisions



**Figure 1.** Illustration of the geographical divisions of the US used in this study.

Division composition		
Division names	Division code	States within the division
New England	NE	Massachusetts, Connecticut, New Hampshire, Maine, Rhode Island and Vermont
Mid-Atlantic	MA	New York, Pennsylvania, New Jersey, Maryland and Delaware
South Atlantic	SA	Florida, Georgia, North Carolina, Virginia, South Carolina and West Virginia
East South Central	ESC	Tennessee, Alabama, Kentucky and Mississippi
West South Central	WSC	Texas, Louisiana, Oklahoma and Arkansas
East North Central	ENC	Illinois, Ohio, Michigan, Indiana and Wisconsin
West North Central	WNC	Missouri, Minnesota, Iowa, Kansas, Nebraska, South Dakota and North Dakota
Mountains	M	Arizona, Colorado, Utah, Nevada, New Mexico, Idaho, Montana and Wyoming
Pacific	P	California, Washington, Oregon, Hawaii and Alaska

**Table 1.** States of the US integrated into 9 divisions. Maryland and Delaware are moved from South Atlantic to Mid-Atlantic.

**Data description.** The flight data comes from the OpenSky Network, which is a non-profit association that provides open access to real-world air traffic control dataset for research purposes<sup>31</sup>. The OpenSky COVID-19 Flight Dataset ([opensky-network.org](https://opensky-network.org)) was made available in April 2020 and is currently updated on a monthly basis, with the purpose of supporting research on the spread of the pandemic and the associated economic impact. This dataset has been used to investigate mobility in the early months of the pandemic<sup>32</sup> as well as the pandemic's effect on economic indicators<sup>33</sup>.

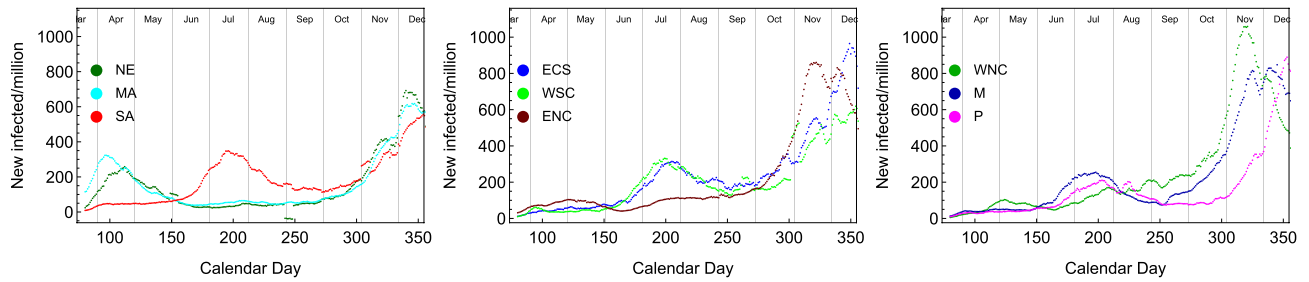
The data provides information about the origin and destination airports as well as the date and time of all flights worldwide. For our analyses we considered domestic flights in the US only. We aggregated the data, to obtain the number of flights between all pairs of airports per day, from the beginning of April until the end of October, 2020. Subsequently, the airports in each state and the number of flights associated with them were combined, to give the number of within and between state flights, on a day to day basis for the whole period.

The number of daily infected cases, which is also used for analysis in this paper, is provided by the open source online repository Opendatasoft ([public.opendatasoft.com/explore/dataset/testing-data-covid19-usa/](https://public.opendatasoft.com/explore/dataset/testing-data-covid19-usa/)).

**Mathematical modeling.** The states within the US have different population and demographic distribution. A state-by-state mathematical modeling, therefore, is challenged by statistical artifacts. For these reasons we group the states following the census divisions ([US Census Bureau](https://www.census.gov)), as summarized in Table 1 and illustrated in Fig. 1. Note, that contrary to the official definitions, we include Maryland and Delaware in Mid-Atlantic instead of South Atlantic. The main reason is that the population of these two states is more connected to states in Mid-Atlantic, as proven by the diffusion timing of the virus.

Building upon our successful understanding of the COVID-19 temporal evolution<sup>34</sup> we apply our framework to the US case. Building on that framework we employ the following eRG set of first order differential equations<sup>15</sup> to describe the time-evolution of the cumulative number of infected cases within the US divisions:

$$\frac{d\alpha_i}{dt} = \gamma_i \alpha_i \left( 1 - \frac{\alpha_i}{a_i} \right) + \sum_{j \neq i} \frac{k_{ij}}{n_{mi}} (e^{\alpha_j - \alpha_i} - 1), \quad (1)$$



**Figure 2.** Weekly new number of cases for all the 9 divisions.

<i>k<sub>ij</sub></i> values (1st wave fits)									
Division code	NE	MA	SA	ESC	WSC	ENC	WNC	M	P
MA	0.72	0	0.0014	0.00075	0.0017	0.0023	0.0005	0.002	0.0053
First wave <i>k<sub>ij</sub></i> values (Flight data, from April 1st to May 31st)									
NE	0	0.72	0.0045	0.00088	0.00087	0.0024	0.00052	0.00067	0.00091
MA	0.72	0	0.019	0.0056	0.0041	0.012	0.0025	0.0031	0.0059
SA	0.0043	0.018	0	0.0085	0.013	0.019	0.0057	0.0050	0.0067
ESC	0.00092	0.0053	0.0093	0	0.0051	0.0068	0.0023	0.0035	0.0065
WSC	0.00095	0.0038	0.014	0.0055	0	0.0092	0.0054	0.011	0.010
ENC	0.0025	0.012	0.018	0.0063	0.0086	0	0.0082	0.0079	0.0099
WNC	0.00038	0.0022	0.0056	0.0019	0.0046	0.0070	0	0.0055	0.0027
M	0.00050	0.0020	0.0042	0.0026	0.011	0.0072	0.0043	0	0.028
P	0.00084	0.0055	0.0063	0.0050	0.010	0.0092	0.0033	0.030	0
Second wave <i>k<sub>ij</sub></i> values (Flight data, from September 1st to October 31st)									
Region-X	0.0066	0.028	0.029	0.013	0.019	0.027	0.014	0.03	0.03
NE	0	0.72	0.0028	0.00046	0.00031	0.0015	0.00026	0.00041	0.00082
MA	0.72	0	0.011	0.002	0.0017	0.0064	0.0013	0.0021	0.0029
SA	0.0026	0.011	0	0.005	0.005	0.0096	0.003	0.0033	0.0035
ESC	0.00041	0.0019	0.0051	0	0.0019	0.0028	0.00087	0.0012	0.0015
WSC	0.00028	0.0015	0.0049	0.0018	0	0.0028	0.0016	0.004	0.0034
ENC	0.0014	0.0062	0.0089	0.0028	0.003	0	0.0039	0.0043	0.0045
WNC	0.00024	0.0013	0.0028	0.0009	0.0017	0.0038	0	0.0028	0.0016
M	0.00032	0.0017	0.0029	0.0011	0.0054	0.004	0.0026	0	0.014
P	0.00074	0.0028	0.0032	0.0014	0.0046	0.0041	0.0018	0.015	0

**Table 2.** Values of the *k<sub>ij</sub>* entries among US divisions. In the top section, the values between Mid-Atlantic (MA) and the other divisions are obtained from fits of the first wave timing. In the central and bottom sections, the complete matrix (except the entries between MA and NE) is obtained using flight data for the first wave (from April 1st to May 31st) and the second (from September 1st to October 31st), respectively.

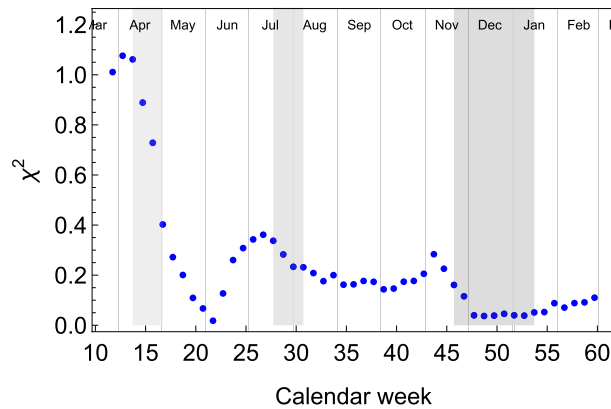
where

$$\alpha_i(t) = \ln \mathcal{I}_i(t), \tag{2}$$

with  $\mathcal{I}_i(t)$  being the cumulative number of infected cases *per million* inhabitants for the division *i* and  $\ln$  indicating its natural logarithm. These equations embody, within a small number of parameters, the pandemic spreading dynamics across coupled regions of the world via the temporal evolution of  $\alpha_i(t)$ . The parameters  $\gamma_i$  and  $a_i$  can be extracted by the data within each single wave. The fit methodology is described in<sup>14,15</sup>.

In the US, it is well known that the COVID-19 pandemic started in NE and MA (mainly in New York City) and then spread to the other divisions. Thus, we define the *US first wave* period from March to the end of August as shown in Fig. 2. In particular, one observes a peak of new infected in NE and MA around April, while for the other divisions the main peak occurs around July. We also observe an initial feature in the latter divisions that we did not attempt to model except for ENC (mostly located in Chicago) and WNC. For the two latter divisions, we considered these as two independent first wave components. The *US second wave* is thus associated with the episode starting in October, 2020.

As a first method, working under the assumption that the US pandemic indeed originated in New York (MA), we first determine the *k<sub>ij</sub>* matrix entries between the division MA and the others. The values are chosen to reasonably reproduce the delay between the main peaks of the first wave in pairs of divisions (C.f. the top section in Table 2). Interestingly, with the exception of NE, the entries of the *k* matrix are comparable to the ones



**Figure 3.** Evolution of the uniformity indicator  $\chi^2$  over time (weekly basis). The shaded bands indicate the period when epidemic peaks are recorded.

we used for Europe<sup>34</sup>. For NE, a large coupling is needed due to the tight connections between the two regions, in particular New York City with the neighbouring states and Massachusetts.

As a second method, we used the flight data to estimate the number of travellers between different divisions, under the assumption that the  $k_{ij}$  matrix entries are proportional to this set of data. To have a realistic matrix for  $k_{ij}$ , we first take the mean number of flights from division  $i$  to division  $j$  during the period from April 1st to May 31st for the first wave, and from September 1st to October 31st for the second wave. Then, we multiply the number of flights by an effective average number of passengers, and normalize it by  $10^6$ , following the definition of  $k_{ij}$ <sup>15</sup>. For the first wave, the optimal average number of passenger is found to be 10, while for the second wave we find an optimal value of 5. Note that these values do not correspond to the actual number of passengers in the flights: in fact, the values of the couplings  $k_{ij}$  also take into account the probability of the passengers to carry the infection as compared to the average in the division of origin. When the value is low it might suggest that the sample of passengers in a flight is less infectious than average, as people with symptoms tend not to travel. Controls at airports may also contribute to this. The key information we extract from the flight data is the relative flux of infections among different divisions.

The results are listed in the middle and bottom sections of Table 2. We keep the same value from the previous fit only for MA-NE. The reason behind this choice is the tight connection between the two divisions, where most of the human mobility is imputable to road transport.

By the end of November, we clearly observe a new rise in the number of infections, signalling the onset of a second wave pandemic in the US (see Fig. 2). Using our framework, we model and then simulate the second wave across the different US divisions.

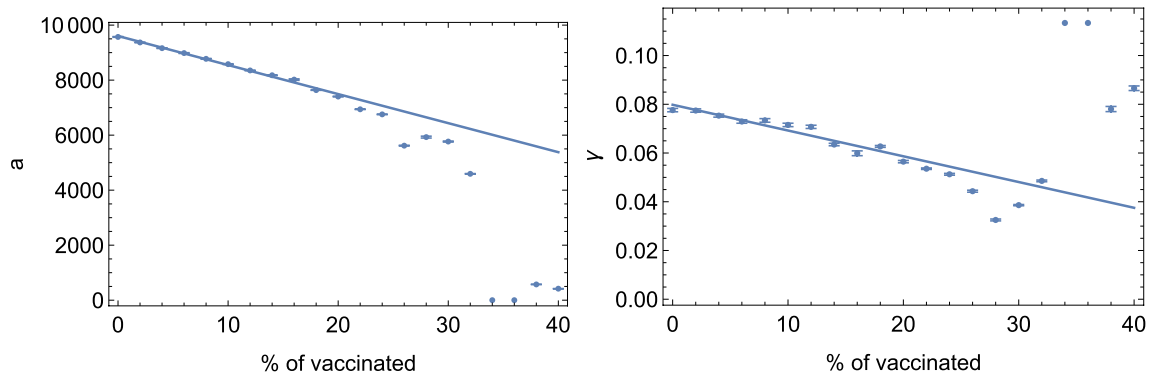
Finally, to check the geographical diffusion of the virus during the various phases of the pandemic in the US, we define an indicator of the uniformity of the new case incidence<sup>18</sup>. This indicator can be defined week by week via a  $\chi^2$ -like variable, given by:

$$\chi^2(t) = \frac{1}{9} \sum_{i=1}^9 \left( \frac{\mathcal{I}_i'(t)}{\langle \mathcal{I}'(t) \rangle} - 1 \right)^2, \quad (3)$$

where  $\mathcal{I}_i'(t)$  is the number of new cases per week in division  $i$  at time  $t$  and  $\langle \mathcal{I}'(t) \rangle$  the mean of the same quantity in the 9 divisions. The parameter  $\chi^2$  quantifies the geographical diffusion of the SARS-CoV-2 virus in the US: the smaller its value, the more uniform the pandemic spread within the whole country. The result is shown in Fig. 3: during the first peak in April (light gray shade), the value of  $\chi^2$  is large, signalling that the epidemic diffusion is localized in a few divisions; during the second peak of the first wave (gray shade), the value has dropped, signalling that the epidemic has been spreading to all divisions. Finally, the data for the ongoing second wave (dark gray shade) shows that  $\chi^2$  is dropping towards zero, as expected for a more diffuse incidence of infections.

**Vaccine deployment and implementation.** Various vaccines have been developed for the COVID-19 pandemic, and their deployment in the US has already started on December 14th (<https://www.washingtonpost.com>). The effect of the immunization due to the vaccine has been studied in the context of compartmental models, like SEIR<sup>30</sup>. In our mathematical model, the simplest and most intuitive effect is a reduction of both the total number of infections during a single wave,  $a_i$ , and/or the effective diffusion rate of the virus  $\gamma_i$ , in each division.

To validate this working hypothesis, and understand how the vaccination of a portion of the population affects the values of  $a$  and  $\gamma$  in the eRG framework, we studied the effect of immunization in a simple percolation model, which has been shown to be in the same class of universality as simple compartmental models<sup>35</sup>. To do so, we set up a Monte-Carlo simulation, consisting of a square grid whose nodes are associated to a susceptible individual. Each node can be in four exclusive states: Susceptible (S), Infected (I), Recovered (R) or Vaccinated (V). At each step in time in the simulation, for each node we generate a random number  $r$  between 0 and 1: if the node is in state S in proximity to a node in state I and  $r < \gamma_*$ , we switch its state to I, else it remains S; if the



**Figure 4.**  $a$  and  $\gamma$  fit parameters versus initial percentage of vaccinated nodes for  $\gamma_* = 0.6$  and  $\epsilon_* = 0.4$ .

node is in state I and  $r < \epsilon_*$ , we switch its state to R, else it remains I; if the node is in state R or V, it will not change. This model reproduces the diffusion of the infection, where  $\gamma_*$  is the infection probability on the lattice and  $\epsilon_*$  is the recovery rate. Finally, we fit the data from the simulation to the solution of a simple eRG equation to extract  $\gamma$  and  $a$ . The vaccination is implemented by setting a random fraction  $R_v$  of nodes to the state V before the simulation starts. The values of  $a$  and  $\gamma$  as a function of the fraction of vaccinations are shown in Fig. 4: we observe that both parameters are reduced by the same percentage as the vaccination up to  $R_v \lesssim 25\%$ . Above this value of vaccinated nodes, the simulation is unstable and the result cannot be trusted. This result, nevertheless, demonstrates that the vaccination reduces both parameters  $a$  and  $\gamma$  proportionally reinforcing our expectation.

In a realistic scenario, the vaccination of the population can only be implemented in a gradual way, so that the total vaccination campaign has a duration in time. We can thereby assume that a fraction  $R_v$  of the population is vaccinated in a time interval  $\Delta_t$ . The rate of vaccinations is therefore  $c = R_v/\Delta_t$ . This implies that the variation in  $\gamma$ , during the time interval from  $t_v$  to  $t_v + \Delta_t$ , is given by:

$$\frac{d\gamma(t)}{dt} = -c \gamma(t_v), \quad (4)$$

where  $\gamma(t_v)$  is the effective infection rate before the start of the vaccination campaign. The solution for the time-dependent effective infection rate is

$$\gamma(t) = \gamma(t_v)[1 - c(t - t_v)], \quad (5)$$

until  $t = t_v + \Delta_t$ , after which  $\gamma$  remains constant again at a reduced value  $\gamma(t_v)(1 - R_v)$ .

To find the variation of  $a(t)$  within the vaccination interval  $t_v$  to  $t_v + \Delta_t$ , we assume that the not-yet-infected individuals are vaccinated at the same rate  $c$  as the total population. Thus, at any given time, the variation in the number of individuals that will be exposed to the infection,  $\mathcal{J}_{\text{exp}}(t) = e^{a(t)}$ , is proportional to the difference  $e^{a(t)} - e^{\alpha(t)}$ . This leads to the following differential equation:

$$\frac{da(t)}{dt} = -c(1 - e^{\alpha(t)-a(t)}) = -c \left(1 - \frac{\mathcal{J}(t)}{\mathcal{J}_{\text{exp}}(t)}\right). \quad (6)$$

This equation needs to be solved in a coupled system with the eRG one. Note that the derivative is zero outside of the time interval  $[t_v, t_v + \Delta_t]$ . In the numerical solutions for the effect of the vaccine, we will add one equation for each  $a_i(t)$ , assuming that the vaccination rate  $c$  is the same in all divisions.

## Results

**Validating the eRG on the first wave data.** The epidemic data (C.f. Fig. 2) shows that the MA division (New York City) was first hit hard by the COVID-19 pandemic, and was followed closely by NE. The other divisions witnessed a comparable peak of new infections 3–4 months later. Note that we are using cases normalized per million to facilitate the comparison between divisions with different population. As a first study, we want to test the eRG equations (1) against the hypothesis that the epidemic has been diffusing from MA to the other divisions. The parameters  $a_i$  and  $\gamma_i$  are fixed by fitting the data, as shown in Table 3. Thus, the timing of the peaks in the divisions is determined by the entries of the  $k_{ij}$  matrix. Determining all 81 entries from the data is not possible, as we only have 9 epidemiological curves. Thus, we assume that only the couplings between the source MA and any other division are responsible. The results of the fits are shown in the top block of Table 2, and will be used as a control benchmark.

Except for  $k_{21}$  that links NE and MA, all the other  $k_{2j}$  are of order  $10^{-3}$ , thus confirming the range we found for the European second wave<sup>34</sup>. The value of  $k_{21}$  is of order unity, which implies that there is a stronger connection between the two divisions. This may be explained by the fact that there exist a significant flow of people between New York City and the neighbouring states (including Massachusetts) in New England. Work commutes and weekend travelling by car explains the required high number of travellers per week. Another interesting feature is the presence of a small peak of infections for ENC and WNC, around March. This feature cannot originate from the MA division, as that would imply a  $k$ -value of order 10, which is clearly unrealistic<sup>15</sup>. The only viable

Division	Code	First wave parameters (fitted)		Second wave parameters	
		$a$	$\gamma$	$a$	$\gamma$
New England	NE	9.397(7)	0.416(7)	11.006	0.214
Mid-Atlantic	MA	9.496(7)	0.516(9)	10.882	0.206
South Atlantic	SA	9.691(5)	0.370(3)	10.885	0.185
East South Central	ESC	9.63(2)	0.331(6)	11.201	0.207
West South Central	WSC	9.720(6)	0.340(3)	10.713	0.213
East North Central	ENC	8.88(3)	0.300(8)	11.074	0.250
West North Central	WNC	8.66(2)	0.342(6)	11.060	0.263
Mountain	M	9.478(9)	0.330(4)	11.089	0.213
Pacific	P	9.33(1)	0.291(5)	11.535	0.171

**Table 3.** Parameters of the eRG model for the first and second wave in the 9 divisions. For the first wave, we report the values from the fit, including the  $1\sigma$  error. For the second wave, the values are chosen to reproduce the current data, adjourned to December 16th.

solution is that the epidemic hit these two divisions from abroad. On the other hand, the second peak observed around August can be explained by the interaction with MA.

The values of  $k_{ij}$  are, in principle, determined by the flow of people between different divisions. Thus, we could use any set of mobility data<sup>36</sup> to estimate the relative numbers of the entries, while the normalization also depends on the effective infection power of the traveling individuals and it can be determined from the data. With the help of mobility data, we can reduce the 81 parameters to a single one. Due to the large distances across divisions, we decided to focus on the flight data, as described in the methodology section. The values of the entries are reported in the middle section of Table 2. Note that for MA-NE we used the same value obtained from the previous fit, as the people's flow is mostly dominated by land movements.

Using this matrix of  $k_{ij}$  to simulate the spread of the first wave across the country, as originating from MA, we obtain the curves in the left panels of Fig. 5. For nearly all divisions, we obtain the correct timing for the peak, with the exception of SA and ESC (for ENC and WNC, the anomaly may be linked to the presence of a mild early peak and the absence of a prominent second peak). The results are more accurate for divisions far from MA, thus validating the method as the diffusion of the virus seems to depend on the people travelling (by air) among divisions. For SA, the predicted curve is substantially anticipated compared to the data: this discrepancy may be explained by the presence of an air hub in Atlanta, GE, so that many of the passengers of flights landing there do not stop in the division but instead take an immediate connecting flight.

**Understanding the second wave.** The US states are currently witnessing a second wave, which is ravaging in all the 9 divisions with comparable intensity. Previous studies in the eRG framework have uncovered two possible origins for an epidemic wave to start: one is the coupling with an external region with a raging epidemic<sup>15</sup>, the second is the instability represented by a strolling phase in between waves<sup>17,18</sup>. We have shown that the former mechanism can account for the peak structure during the first wave.

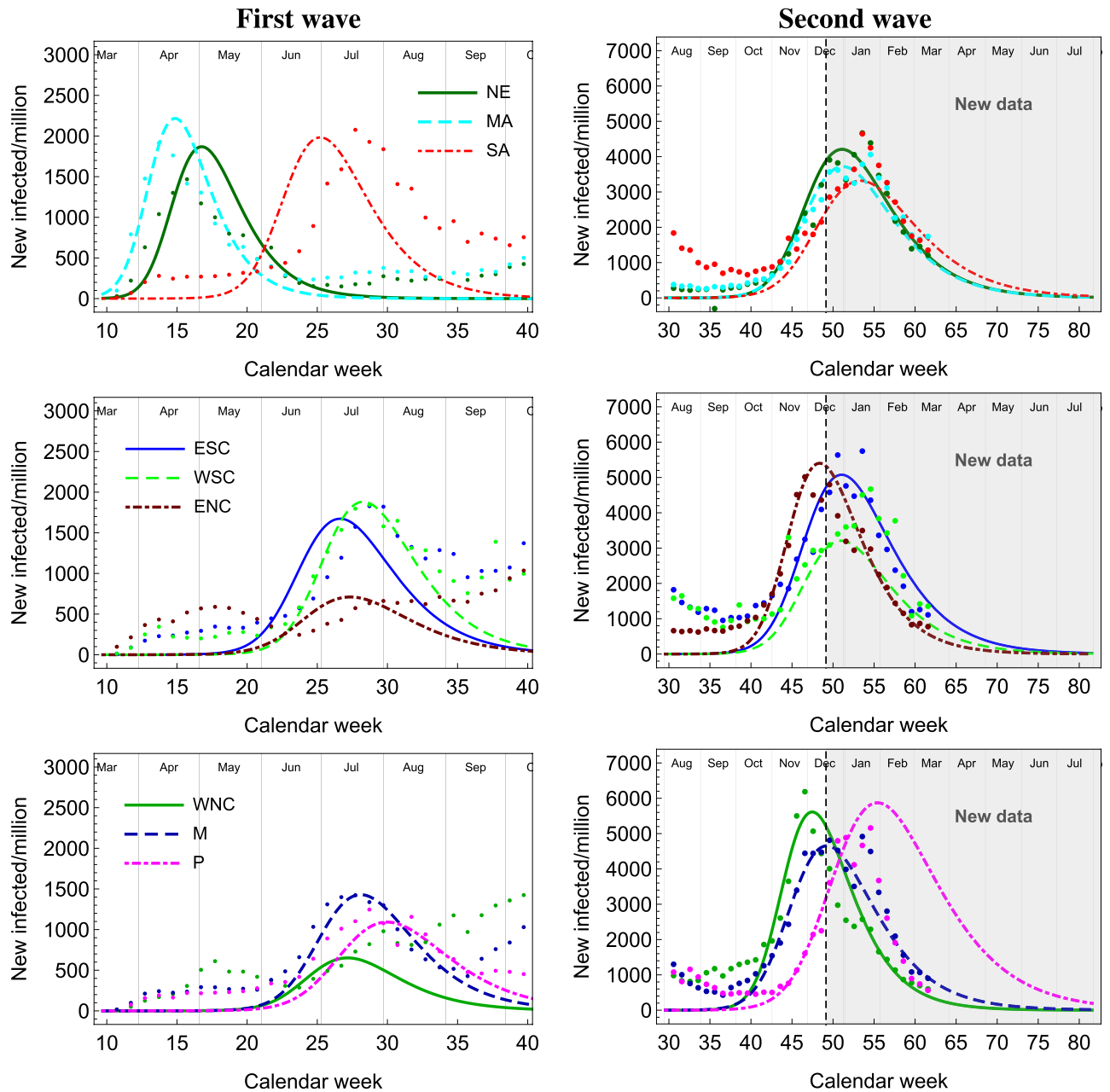
As a first step, we will try to use the same method to understand the second wave. Since travelling to the US from abroad has been strongly reduced and regulated, we will consider the divisions that witnessed a peak in July–August as source for the second wave. To this purpose, we define a Region-X<sup>15</sup> as an average sum of all the divisions with a pandemic peak occurring in the July–August period. The parameters are chosen to reproduce the number of cases in the totality of the relevant 7 divisions (SA, ESC, NSC, ENC, WNC, M and P) normalized by the total population. For each division, we optimized  $a_i$  and  $\gamma_i$  to reproduce the current data adjourned at December 16th (C.f. Table 3). For the couplings  $k_{ij}$  we use the flight data, except for the usual MA-NE couplings (C.f. bottom section of Table 2). Finally, the  $k_{0j}$  connecting the 9 divisions to the source Region-X are computed by summing the  $k$  entries between the division  $j$  and the 7 divisions used to model Region-X (also derived from flight data).

The results of the eRG equations are shown in the right panels of Fig. 5, showing a good agreement. The values of the  $k_{0j}$  of the Region-X are one order of magnitude larger than the others. This fact can be interpreted by the presence of hotspots in each division which also contribute significantly to the new wave. In other words, traveling among divisions cannot be the only responsible factor for the onset of the second wave in the US. This hypothesis can also be validated by studying the uniformity of the distribution of the new infections in various states during the three peaks, as shown in Fig. 3. Comparing the three peak regions, we see that the uniformity indicator is systematically decreasing, thus indicating a more geographically uniform presence of the virus.

It is also interesting to notice that the value of  $\gamma_i$  for the second wave is systematically smaller than the infection rate during the first wave. This is in agreement with the results we found in Ref.<sup>17,18</sup>, where we modelled the multi-wave structure of the pandemic via an instability inside each region. The result of this simple analysis supports the hypothesis that the virus is now endemic for all states in the US, thus a multi-wave pattern will continue to emerge. Traveling among states (or divisions) is less relevant at this stage.

The result of our eRG analysis shows that the current wave will end in March–April 2021. Note, however, that we have not taken into account the potential disastrous effect of the Christmas and New Year holidays, which

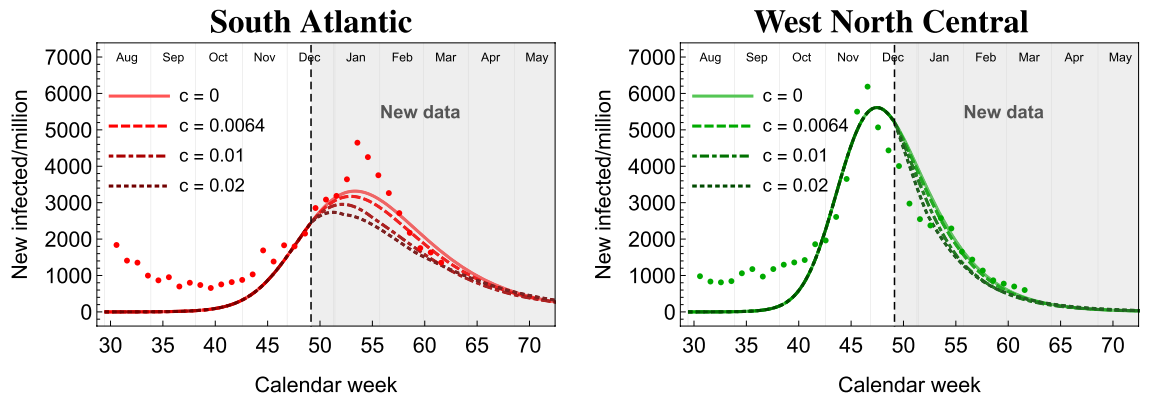




**Figure 5.** Simulation of the spread of the first wave (left plots) and the second wave (right plots) using flight-data-derived kappa matrix. For the first wave, MA is used as a seed region, while for the second wave a combination of the first waves among divisions acts as the seed region (Region-X). The vertical dashed lines in the right column plots mark the date when the simulation was done. The data points in the grayed region where the data were not used to tune the eRG solutions.

could lead to an increase in the infection rates. In some divisions there is an increase at the end of November, which can be attributed to the Thanksgiving holiday.

**Effect of the current vaccination strategy.** Following the development of multiple vaccines for the SARS-CoV-2 virus (<https://www.nature.com/articles/d41586-020-03626-1>), vaccination campaigns have started in many countries, including the US. This will influence the development of the current wave, and help in curbing the future ones. The vaccination campaign started on December 14th in the US (<https://www.washingtonpost.com>). We also know that the US has purchased 100 million doses from Pfizer (plus an additional 100 million from Moderna) (<https://www.forbes.com>), so that at least 20% of the population may be vaccinated in this first campaign. As of December 28, 0.64% of the total population has been vaccinated (<https://ourworldindata.org>) in a little over 1 week, thus in our study we will use this as a benchmark weekly rate. The data listed above defines our starting benchmark for the current vaccination campaign.



**Figure 6.** Evolution of the number of infections without vaccination ( $c = 0$ ) and with a vaccination rate of 0.64%/week, 1%/week and 2%/week starting on December 14th and stopping at 20% of the population vaccinated. We show the results for two sample divisions: South Atlantic and West North Central.

Division	Code	Percentage of the population vaccinated		Vaccination rate slope $\mu$	
		Partial vaccination	Full vaccination	Partial vaccination	Full vaccination
New England	NE	30.1	16.4	0.00450	0.00641
Mid-Atlantic	MA	27.8	14.1	0.00415	0.00554
South Atlantic	SA	24.3	13.5	0.00363	0.00530
East South Central	ESC	23.1	12.6	0.00345	0.00492
West South Central	WSC	23.0	11.9	0.00344	0.00467
East North Central	ENC	25.6	14.7	0.00382	0.00575
West North Central	WNC	26.3	14.9	0.00393	0.00584
Mountain	M	25.7	14.8	0.00386	0.00579
Pacific	P	26.6	14.0	0.00397	0.00548

**Table 4.** Percentage of the population vaccinated with at least one dose and with two doses in each division as of the date of 24th of march 2021.

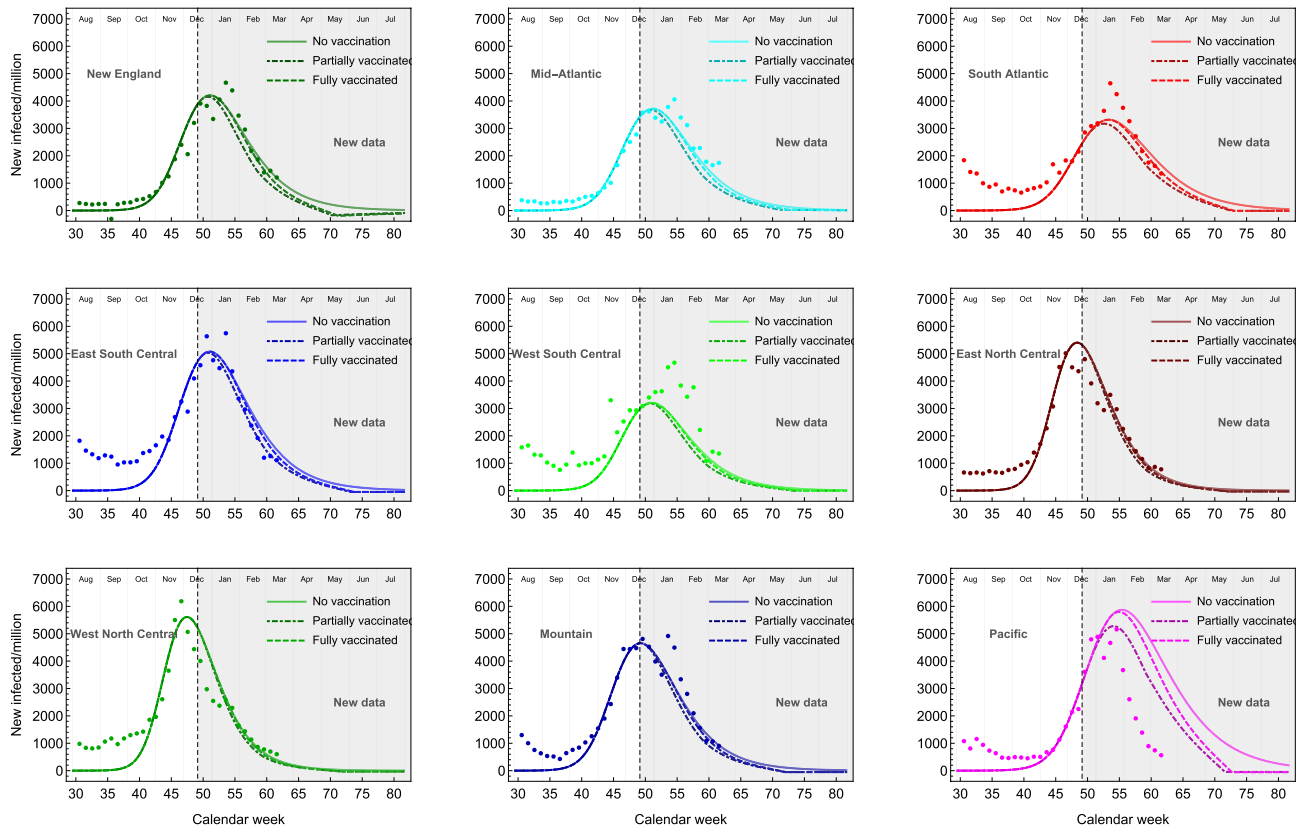
To study the effect of the vaccinations, we have solved the eRG equations for the second wave, with the addition of the reduction of  $a_i$  and  $\gamma_i$ , as detailed in the methodology section. We show the result for two sample divisions in Fig. 6 (dashed curves) as compared to the same solutions without vaccines (solid curves). A vaccination at a 0.64% rate per week does not affect the peak of new infections. As a reference, we also increased the vaccination rates to 1% and 2%: in these cases, an important flattening of the epidemic curve can be observed for SA, where the vaccination started early compared to the peak of infections. This situation may be realized, as the vaccine is being administered to the population that is more at risk of being infected by the virus. In the other extreme case, represented by WNC, the vaccine is ineffective in changing the current wave because the peak has already been attained before the vaccination campaign started.

Our results confirm that the current vaccination strategy, which is performed during a peak episode, is not effective to substantially slow down the spread of the virus. On the other hand, the effectiveness for future waves is not a question. It would be, in fact, very efficient to be able to administer the vaccine to a larger portion of the population before the start of the next wave.

**Update of the vaccination to the first quarter 2021.** As shown in the right column in Fig. 5, our simulation of the second wave, done in mid December 2020, reproduces very well the epidemiological data up to March 17, 2021. The only exception is Pacific, which has seen a sharper drop in the number of new infections. Furthermore, one can clearly see a rebound in January that can be accredited to the Christmas holidays. Nevertheless, this small effect does not have a significant impact on the agreement of our prediction with the data.

In the first quarter of 2021, the vaccination campaign has also taken off steadily, with nearly a quarter of the US population having received at least one shot of vaccine. Furthermore, since February 27 the FDA has authorised the use of the Janssen mono-dose vaccine (<https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/janssen-covid-19-vaccine>), which is now being administered together with the two-dose Pfizer and Moderna vaccines. The data show that the rate of vaccinations has been increasing approximately linearly with time, thus we updated the prediction to take into account a vaccination fraction  $c(t)$  growing linearly with time:





**Figure 7.** Results of the eRG solutions for the second wave with a vaccination campaign based on the data. Here, we consider a vaccination rate linearly increasing in time, with slopes given in Table 4. The eRG parameters are the same used for Fig. 5, based on data until December 28, 2020.

$$c(t) = u t, \tag{7}$$

where the numerical values for each division are shown in Table 4.

The new results are shown in Fig. 7 for the 9 divisions. We consider both people that received at least one shot (partial vaccination, in dash-dotted lines) and fully vaccinated ones (dashed lines), with an offset of 4 weeks between the beginning of the two. We consider them as two extreme cases, defining a systematic error in our account of vaccinations. In all divisions, the effect is minor, as the vaccination campaign has started too close to the peak of the second wave. The only exception is Pacific, where taking into account vaccinations substantially improves the agreement with the data.

The updated results confirm that a vaccination campaign operating during a wave will not significantly affect the timing and height of the peak. Social distancing and containment measures remain necessary. Conversely, vaccinating a large portion of the population will certainly curb the eventual next wave.

### Discussion

In this paper we employ the epidemic Renormalization Group (eRG) framework in order to understand, reproduce and predict the diffusion of the COVID-19 pandemic across the US as well as the effect of vaccination strategies. By using flight data, we are able to see the changes in mobility across the divisions, and observe how these changes affect the spread of the virus. Furthermore, we show that the impact of the vaccination campaign on the current wave of the pandemic in the US is marginal. Based on that, the importance of social distancing is still relevant. Furthermore, we demonstrate that the current wave is due to the endemic diffusion of the virus. Therefore, building upon our previous results<sup>18</sup>, in order to control the next pandemic wave the number of daily new cases per million must be around or less than 10–20 during the next inter-wave period. This conclusion is further corroborated in Ref.<sup>37</sup> for Europe.

We learnt that the number of infected individuals in the current wave are not affected measurably by the vaccination campaign. However, it is foreseeable that it will impact specific compartments such as the overall number of deceased individuals. Our study included an immunization rate between 0.64 to 2% of the total population each week. We also updated the results with the actual rates of vaccination in the different divisions, as of March 24, 2020. The results of our eRG model agree remarkably well with the new data from December 28, 2020, to March 17, 2021. To curb the current and the next waves, our results indisputably show that vaccinations alone are not enough and strict social distancing measures are required until sufficient immunity is achieved.

Our results should be employed by governments and decision makers to implement local and global measures and, most importantly, the results of this paper can be used as a foundation for vaccination campaign strategies for governments.

Given that pandemics are recurrent events, our results go beyond COVID-19 and are universally applicable. What we have seen in the data for the US is that it started in New York and, from there, it diffused to the rest of the country. It is, therefore, important to contain future pandemics at an early stage.

## Data and code availability

All the data used in this work is from open source repositories, cited in the main text. The Wolfram Mathematica code will be provided upon request.

Received: 13 January 2021; Accepted: 29 April 2021

Published online: 26 May 2021

## References

- Perc, M., Gorišek Miksić, N., Slavinec, M. & Stožer, A. Forecasting COVID-19. *Front. Phys.* **8**, 127. <https://doi.org/10.3389/fphy.2020.00127> (2020).
- Hâncean, M.-G., Perc, M. & Juergen, L. Early spread of COVID-19 in Romania: Imported cases from Italy and human-to-human transmission networks. *R. Soc. Open Sci.* **7**, 200780. <https://doi.org/10.1098/rsos.200780> (2020).
- Zhou, T. *et al.* Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *J. Evid. Based Med.* **13**, 3–7. <https://doi.org/10.1098/rspa.1927.0118> (2020).
- Prem, K. *et al.* The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**(5), E261–E270. [https://doi.org/10.1016/S2468-2667\(20\)30073-6](https://doi.org/10.1016/S2468-2667(20)30073-6) (2020).
- Scala, A. *et al.* Time, space and social interactions: Exit mechanisms for the COVID-19 epidemics. *Sci. Rep.* **10**, 13764. <https://doi.org/10.1038/s41598-020-70631-9> (2020).
- Friston, K. J. *et al.* Second waves, social distancing, and the spread of COVID-19 across America. *Wellcome Open Res.* **5**, 103. <https://doi.org/10.12688/wellcomeopenres.15986.3> (2021).
- Sonnino, G., Mora, F. & Nardone, P. A stochastic compartmental model for COVID-19 (2020). [arxiv: 2012.01869](https://arxiv.org/abs/2012.01869).
- Abou-Ismaïl, A. Compartmental models of the COVID-19 pandemic for physicians and physician-scientists. *SN Compr. Clin. Med.* <https://doi.org/10.1007/s42399-020-00330-z> (2020).
- Kermack, W. O., McKendrick, A. & Walker, G. T. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* **115**, 700–721. <https://doi.org/10.1098/rspa.1927.0118> (1927).
- Zhan, X.-X. *et al.* Coupling dynamics of epidemic spreading and information diffusion on complex networks. *Appl. Math. Comput.* **332**, 437–448. <https://doi.org/10.1016/j.amc.2018.03.050> (2018).
- Perc, M. *et al.* Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51. <https://doi.org/10.1016/j.physrep.2017.05.004> (2017).
- Wang, Z., Andrews, M. A., Wu, Z.-X., Wang, L. & Bauch, C. T. Coupled disease-behavior dynamics on complex networks: A review. *Phys. Life Rev.* **15**, 1–29. <https://doi.org/10.1016/j.plrev.2015.07.006> (2015).
- Scudellari, M. How the pandemic might play out in 2021 and beyond. *Nature* **584**, 22–25. <https://doi.org/10.1038/d41586-020-02278-5> (2020).
- Della Morte, M., Orlando, D. & Sannino, F. Renormalization group approach to pandemics: The COVID-19 case. *Front. Phys.* **8**, 144. <https://doi.org/10.3389/fphy.2020.00144> (2020).
- Cacciapaglia, G. & Sannino, F. Interplay of social distancing and border restrictions for pandemics (COVID-19) via the epidemic Renormalisation Group framework. *Sci. Rep.* **10**, 15828. <https://doi.org/10.1038/s41598-020-72175-4> (2020). [arxiv: 2005.04956](https://arxiv.org/abs/2005.04956).
- Della Morte, M. & Sannino, F. Renormalisation Group approach to pandemics as a time-dependent SIR model. *Front. Phys.* **8**, 583. <https://doi.org/10.3389/fphy.2020.00144> (2021).
- Cacciapaglia, G. & Sannino, F. Evidence for complex fixed points in pandemic data. <https://doi.org/10.21203/rs.3.rs-70238/v1> (2020). [arxiv: 2009.08861](https://arxiv.org/abs/2009.08861).
- Cacciapaglia, G., Cot, C. & Sannino, F. Multiwave pandemic dynamics explained: How to tame the next wave of infectious diseases. *Sci. Rep.* **11**, 6638. <https://doi.org/10.1038/s41598-021-85875-2> (2021).
- Taubenberger, J. K. & Morens, D. M. 1918 influenza: The mother of all pandemics. *Rev. Biomed.* **17**(1), 69–79 (2006).
- Wilson, K. G. Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture. *Phys. Rev. B* **4**, 3174–3183. <https://doi.org/10.1103/PhysRevB.4.3174> (1971).
- Wilson, K. G. Renormalization group and critical phenomena. 2. Phase space cell analysis of critical behavior. *Phys. Rev. B* **4**, 3184–3205. <https://doi.org/10.1103/PhysRevB.4.3184> (1971).
- Li, L. *et al.* Analysis of transmission dynamics for Zika virus on networks. *Appl. Math. Comput.* **347**, 566–577. <https://doi.org/10.1016/j.amc.2018.11.042> (2019).
- Wang, Z. *et al.* Statistical physics of vaccination. *Phys. Rep.* **664**, 1–113. <https://doi.org/10.1016/j.physrep.2016.10.006> (2016).
- Danby, J. M. A. *Computing applications to differential equations modelling in the physical and social sciences* (Reston Publishing Company, Reston\*\*\*, 1985).
- Brauer, F. Early estimates of epidemic final sizes. *J. Biol. Dyn.* **13**, 23–30. <https://doi.org/10.1080/17513758.2018.1469792> (2019).
- Miller, J. C. A note on the derivation of epidemic final sizes. *Bull. Math. Biol.* **74**, 2125–2141. <https://doi.org/10.1007/s11538-012-9749-6> (2012).
- Murray, J. D. *Mathematical Biology. Interdisciplinary Applied Mathematics* 3rd edn. (Springer, New York\*\*\*, 2002).
- Fishman, D., Kho, E. & Tuite, A. Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model. *PLoS Curr. Outbreaks* **6**, (2014).
- Pell, B., Kuang, Y., Viboud, C. & Chowell, G. Using phenomenological models for forecasting the Ebola challenge. *Epidemics* **22**, 62–70. <https://doi.org/10.1016/j.epidem.2016.11.002> (2018) (**The RAPIDD Ebola Forecasting Challenge**).
- Paltiel, A. D., Schwartz, J. L., Zheng, A. & Walensky, R. P. Clinical outcomes of a COVID-19 vaccine: Implementation over efficacy. *Health Affairs* **Vo.40**, No.1 (2020).
- Schäfer, M., Strohmeier, M., Lenders, V., Martinovic, I. & Wilhelm, M. Bringing up opensky: A large-scale ads-b sensor network for research. In *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, 83–94 (IEEE, 2014).
- Islind, A. S., Óskarsdóttir, M. & Steingrimsdóttir, H. Changes in mobility patterns in Europe during the COVID-19 pandemic: Novel insights using open source data (2020). [arxiv: 2008.10505](https://arxiv.org/abs/2008.10505).
- Bank of England. Monetary policy report and interim financial stability report—may 2020. <https://www.bankofengland.co.uk/report/2020/monetary-policy-report-financial-stability-report-may-2020> (2020).
- Cacciapaglia, G., Cot, C. & Sannino, F. Second wave COVID-19 pandemics in Europe: A temporal playbook. *Sci. Rep.* **10**, 15514. <https://doi.org/10.1038/s41598-020-72611-5> (2020). [arxiv: 2007.13100](https://arxiv.org/abs/2007.13100).

35. Cardy, J. L. & Grassberger, P. Epidemic models and percolation. *J. Phys. A Math. Gen.* **18**, L267–L271. <https://doi.org/10.1088/0305-4470/18/6/001> (1985).
36. Yang, C. *et al.* Taking the pulse of COVID-19: A spatiotemporal perspective. *Int. J. Digit. Earth* **13**, 1186–1211. <https://doi.org/10.1080/17538947.2020.1809723> (2020).
37. Priesemann, V. *et al.* Calling for pan-european commitment for rapid and sustained reduction in SARS-CoV2 infections. *Lancet* **397**(10269), 92–93 (2020).

### Author contributions

This work has been designed, analysed and performed conjointly and equally by the authors CC, GC, ASI, MO and FS, who have equally contributed to the writing of the article. ASI and MO have extracted and processed the data from flights; CC has worked on the numerical results from the eRG equations and analysed the epidemiological data.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to C.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021