A comercial airline network model for understanding the chikungunya spread in the Caribbean.

Abstract
In an increasingly globalised and more interconnected world the emergence and re-emergence of infectious diseases has become a non-negligible threat and a real concern for modern societies. Infectious diseases have nowadays the capacity to travel longer and reach remote locations in relatively shorter times due to the human's growing ability for mobility as transport has became accessible to the masses. This onward mobility capacity spans geographical scales from local to global. Here, a commercial airline database has been used to set a human mobility network model. This network model has in turn been coupled to a SEIR epidemic model in order to simulate the first stages of the CHIKV spread in the subpopulations that started in the end of 2013 in the Caribbean region. The data-fitted model explores the role of the topology of the particular flight network of the Caribbean -and thus the mobility of infectious hosts- on the initial spatial and temporal patterns of incidence of CHIKV in the set of the firstly affected islands and mainlands of the outbreak. Several scenarios has been investigated for such purpose. Among these case studies we assess a number of rearrangements of the observed network topology and their effect on the spatio-temporal dynamics of the outbreaks, and we make a sensitivity analysis of the model by evaluating the epidemics implications of outbreaks starting in other places instead of Saint Martin island. Finally, simulations of the data-fitted model of a simple intervention action demonstrate the appearance of a limiting time-horizon where any level control efforts later than this time have null effect to lessen the epidemic severity. The model reveals that, for interventions strategies to be useful in controlling outbreaks of emergent infectious diseases in naive populations such as CHIKV in the Americas in 2013, the understanding of the dynamics of the initial stages is crucial, especially when occurring in heterogeneous loosely connected host populations.

Introduction
Chikungunya fever is an acute febrile illness that despite re-emerging in 2004, it is only since its introduction in the Caribbean in late 2013, that it underwent an unprecedented and unpredictable spread in the form of new massive suspected infections in a naive territory. The explosive nature of its spatial propagation since the first confirmed case was reported in the island of St. Martin and the recent appearance of Zika virus in the region, together with its more benign symptoms have somewhat 'eclipsed' its true societal relevance. However, an epidemic that affected 23 countries in the first 6 months, 35 countries in the first 12 months [5] and that in February 2016 already had resulted in more than 1,7 million new infections in the Americas in 45 countries [6], necessary deserves to be considered a medical problem of global concern. Chikungunya virus (CHIKV) was first recognized as a human pathogen during the 1950s in Africa, and since then, cases have been identified in many countries in Africa and Asia [13, 28]. CHIKV likely originated in Central/East Africa, where the virus has been found to circulate in a sylvatic cycle between 1forest dwelling Aedes species mosquitoes and nonhuman primates [23]. Transmitted by the bite of an infected Aedes mosquito, sporadic human cases occur in Africa, but large human outbreaks are infrequent. Instead it easily propagated across large fractions of the population in the American continent due to the ubiquitous nature of its arthropod vectors (mainly Aedes aegypti and Aedes albopictus mosquitoes) in and around highly-populated urban centres. Genetic change likely contributed to the magnitude and distribution of outbreaks, as autochthonous transmission can take place through more than one species such as in La Reunion,

where both Ae. Aegypti and Ae. Albopictus were present[33, 34].

Of concern are the high level of viremia detected in humans[16, 32]) together to its mild symptoms (e.g. during early epidemics rare but serious complications of the infection are noted, which only results in serious medical conditions in a very small proportion of cases[3]. Former factors result in a majority of asymptomatic infections that unnoticed to the public health system, easily and in an uncontrolled manner may rapidly propagate to uninfected areas in islands and villages where the mosquito population is abundant and sanitation levels are not optimal [8]. Clear examples in island settings were the high attack rates in the 2004 outbreak in the Lamu island with more than a 70% of the total population infected[29], the 2005 outbreak in the Comoros with a 63% of the population being infected [30] and in La Reunion island [7] with over 260,000 cases totally occurring. Similarly, in small rural villages like Trapeang Roka in Cambodia the basic reproduction number in the first few weeks of the epidemics rose to values as high as 6 [27].

Major aspects yet remain unknown to the current outbreak in the Caribbean, namely why this particular Central/East African strain was so effective in spreading. In the recent past, attempts to model the movement of individuals as ways to understand/approach the rapid propagation of these infectious diseases have been numerous and approaches diverse (e.g. cellular automata [11], networks [19]), individual-based models [1], and metapopulations [15]. However, few specific models are available that incorporate ecological, entomological and virological factors as a way to predict future disease outbreaks. Spatially-explicit models are necessary to evaluate the efficacy of movement controls [10, 26]. Instead, models that ignore spatial structure can lead to inaccuracy in the prediction of population dynamics [9] unless local spatial heterogeneity is not relevant at the scale of the study. For spatial spread, both distributed contacts and distributed-infective models have been used with success. Here we use a similar approach to that in contact networks to study disease transmission in human populations [18, 19, 25]. Furthermore, despite to the existence of a variety of studies that attempt to explain the complexities of epidemic dynamics when spreading in heterogeneous space (some reference classic examples here), not many of them have produced operational models that could be relatively easy to implement and put to use for public health decision and policy makers.

In this study we model and simulate the first 6 months of the expansion of CHIKV in the Caribbean in 2013 and try to understand what the main factors are in determining how the propagation initially takes place. We then seek to trace the disease at the times when both primary and secondary infections where predominant but only during the first occurrence of the outbreak in each island and leave for future studies the recurrent infection outbreaks and waves that subsequently followed. While the former might be primarily more related to the movement of individuals, the latter would be much more a complex matter determined by the structure of the population, herd immunity locally acquired to the first wave, and local geographic/topography and social and economic determinants. Also because during the first wave, misrecognition for other co-circulating diseases (e.g. mild dengue cases, and more recently Zika) should not be as frequent as when diseases with similar symptoms are established in a population and herd immunity and co-infections play a confounding role. CHIKV infections are often confused with dengue viral infections, because both diseases can present with high temperatures and myalgias in people living in or returning from tropical areas [24]. In addition, both viruses are transmitted by the same species of mosquitoes and may co-circulate, leading to dual infections and concurrent epidemics [18, 20, 20]. And finally because symptoms for the disease might be confounding new and old infections, as they can extend long in the future after the initial infection takes place, with up to 64% of patients

with Chikungunya fever reporting joint stiffness and/or pain 1 year after the initial infection, and 12% still reported symptoms 3–5 years later [2, 4, 31].

The persistence of cases of infection in the American continent during the next years will be presumably attributable to vast numbers of immunologically naive people along the continent, who help sustain viral transmission. Instead, this is in stark contrast to Caribbean islands with more limited populations, which normally would not report cases after the epidemic swept through and most likely because of the development of herd immunity. Maintenance of Chikungunya in the American continent, however, may act as a metapopulation process feeding the Caribbean region in the future. It is therefore of extreme relevance, to focus on the very first stages of the epidemic propagation as containment at much later and more developed stages may be absolutely impractical, not to speak about the costly and often ineffective mosquito eradication strategies.

Additionally, this work delve into the feasibility of developing powerful operational models for emergence on vector borne infections based on simple network mobility patterns that can be obtain from airport statistics. The information gained should/could therefore be used as guidance and serve public health officials to more effectively intervene trying to timely/promptly mitigate its effects and better contain/halt its propagation to naive areas.

Results

As described in the methods section, the epidemic network model (1) has been parameterized and fitted by a recursive least squares method to the chikungunya aggregated case incidence data reported by the Project Tycho Data for Health [21] for the initial 203 days of the outbreak in the Caribbean that started towards the end of 2013 in Saint Martin Island (first date of a reported case coming out of locally transmitted chikunguna is reported as 20 of October 2013 by Project Tycho).

The Figure (2–3) shows the reported aggregated data (red line) for the first 203 days of the chikungnya epidemics in the Caribbean in 2013 in the top left panel, and the data-fitted numerical solution (blue curve) of the SEIR-network model (Eq. 1 in the method section) is shown in the bottom left panel. Their respective spectra are plotted in the right panels (please Xavi comment here a bit about the spectra).

The Figure (2–3) shows that the network-SEIR model captures the jagged shape of the epidemic curve exhibited by the observed time series of aggregated observed chikungunya cases. The multi-peaked pattern of the aggregated observed incidence is made of various local single peak epidemic curves at the several islands locations of the region. Since these local peaks are occurring at different times when they are are aggregated they render the time series shape the jagged serrate aspect (Fig 2–3 top left). This can be observed when plotting together the individual observed time series for locations (not shown here). The multi-peak epidemic curve reveals then the differential timely onset of the outbreaks in the different locations in the region implying a spatio-temporal structure of the chikungunya dynamics. Similarly, the jagged epidemic curve of the model solution of the aggregated nodes (Figure 2–3 bottom left panel)is the product of many single peak models solution epidemics at each node of the network that represents the observed locations. Therefore, the model is capturing the spatio-temporal structure of the infection spread observed in the real data. The Figure (4) shows the individuals epidemic nodes model solutions revealing the pattern of the epidemic wave.

The table (1) shows the model parameter values fitted for the first 203 days of observed cases. These values have been computed by an optimization process

described in the methods section and have been used to simulate the scenarios and intervention plots in this article.

Parameter	Definition	Fitted Value
β	Transmission rate	1.21 day^{-1}
σ	Incubation period	0.35 day^{-1}
γ	Recovery rate	0.57 day^{-1}

Table 1: Parameters of the epidemic model estimated by least squares form the observed incidence series.

The fitted SEIR network model is used to assess the role of the topology of the network of connections by checking four different scenarios where we change the original topology of the network in different ways, and the initial starting node of the infection as well. We also look at a simple intervention scenario where the epidemic severity resulting of cutting off mobility at several degrees as well as cutting off at different times since detection of first infectious is measured.

Epidemic scenarios

We hypothesize that in an epidemic setting where a contagious disease is invading naïve population which is spatially structured, such as the 2013 chikungunya outbreak in the Caribbean region, the geometry of the mobility pattern of the host individuals plays a definite role in how the geographical spread of the disease occurs as the host indiviuals are thought as the main carriers of the disease to remote locations. This hypothesis may be extended to similar settings where nearly distinct subpopulations of hosts that are loosely connected by a complex migration or mobility patterns and have not been exposed previously to the infection. For this scenario we conjecture that the topology of mobility is the most important factor for the disease dynamics. This can certainly be a general case for a number of vector-borne and host-to-host infectious diseases occurring from mid to regional to global geographic scales.

Four scenarios have been then investigated to assess the role of the network topology in the simulated epidemic wave. The models scenarios are described as following:

a) Epidemic scenario one: Network model with randomized connections of the initial infected node

In this scenario we have performed simulations of the fitted network SEIR model (Eqs. 1) where the connections of the initial infected node (Saint Martin) has been randomized. That is, the original links that represent flight connections of Saint Martin have been randomly changed along with it corresponding strengths -which are parameterized as proxies porportional to the mobility rate of individuals between any two nodes-. The number of links of the original observed network is preserved, however. Also, the rest of the connections of the network are kept the same as the original model based on the real data flight connections. The left-down plot of Fig. (5-6) shows 5000 random simulations of aggregated infected individuals in which, at each single simulation, the original six connections of Saint Martin has been randomly rewired to any six of the other nodes in the network. The solid black curve corresponds to the model solution for the number of aggregated infected individuals of the original

network model with an initial infected individual in Saint Martin.

The epidemic curves are plotted up to their global maximum, as we for assessing the effect of the different epidemic scenarios we use the *epidemic severity* measured as a function of a) number of infected at the global epidemic peak; b) time to epidemic peak (related to epidemic duration); and c) accumulated infecteds at the global epidemic peak.

The simulations show that disturbing the network by changing the connections of the initial node randomly, even though keeping that number of connections and preserving the rest of the network intact, can have a profound impact on the course the epidemic wave can take as shown by the many different light-red epidemic curves in the left-down plot of Figure (5-6). The infection can go over many several paths producing a variety of epidemic geographic diffusion patterns due to a seemingly small disturbance in the network structure, i.e. the links of the initial infected node. It is important to remark that even though each simulation is random in the sense that the links for the initial infected node are randomly rearranged for each simulation, the SEIR model itself is deterministic at the level of node/local population, so each curve is deterministic and represent a particular path trajectory of the model solution of the spread of the infection across the network nodes. Hence, since at each single simulation, only 6 links of the initial Saint Martin node are randomly reconnected to the rest $30 - 1 = 29$ nodes and the rest of the network is preserved unaltered. One can easily calculate the number of possible epidemic waves or trajectories as the combinatoric number $\binom{29}{6} = \frac{29!}{29!(29-6)!}$ that is, there are 475020 possible combinations and therefore so different epidemic waves. Due to computational limitations, we have randomly chosen and simulated a random sample of 5000 of this possible trajectories in the hope that this is a representative sample of all possible combinations. The simulated sample represents approximately 1.1% of all possible trajectories.

We notice that the epidemic duration of the outbreak produced by the model based on the observed flight connections lies towards the tail of the simulation distributions of this scenario (i.e. it lies in the longer duration part of the simulations). This could imply that the actual observed epidemic is atypically long if compared to other possible epidemics that could be produced by relatively small random changes in the network topology of the initial infected node, in our case the connections of one node alone, Saint Martin. In other words, the epidemic dynamics can be sensitive to small changes and perturbations in the topology of the network mobility.

b) Epidemic scenario two: Network model with initial infected node connected to all other nodes

In this second model scenario we set the initial node, Saint Martin, to have a connection to all the other nodes in the network while leaving the rest of the connection structure like the original topology based on the real flights. Link strengths are assigned to each of these new connections (from Saint Martin to each of the others) by randomly picking observed strengths values of the actual network. This scenario serves as a null benchmark scenario since by connecting the initial infected node to all other nodes we essentially eliminate the spatial structure and render the model into a quasi complete mixed homogeneous SEIR system where individuals in any node have very similar encounter rates with individuals in any other node. We do not show a plot of this benchmark scenario here but the plot looks unmistakably very much like a regular homogeneous SEIR single peak solution. All 5000 trajectories seems to be nearly identically single-peak shaped and appear as overlapped when plotted (not show here).

In this scenario we randomly vary at each single simulation only the link strengths and keep the the topology constant, therefore the observation that all solution trajectories seemingly lie on a single curve suggests that the link geometry –and not the distribution of link strengths– is the main actor in determining the spatio-temporal unfolding of the disease in the subpopulations of the nodes. Hence, this scenario serves as (i) a benchmark test to assess the role of the topology in the dynamics, and (ii) to show that the strength of the link connections seems to be irrelevant at least at the scales observed in the networks of flights reported for the Caribbean in 2013-14, and/or for the dynamics of the initial stages of an outbreak unfolding in a nearly discontinued set of naive populations.

If we consider that the network topology of host mobility is important to understand the outbreak unfolding in time and space, and that the overall spread pattern is related, among other topological characteristics, with the path length* the contagion chain has to go through, then connecting the first infected node to each of every other node in the network would dramatically change the current network link geometry by reducing the shortest average path length**. The observed network based on the reported flight connection has an average (shortest) path length, $l_G^{obs} = 2.19$, whereas the average path length of the network of the this benchmark scenario is $l_G^{scen2} = 1.754$, which indicates that effectively the average shortest path has been decreases by connecting Saint Martin to all other nodes, which is expected. This makes the contact structure of the model's transmission kernel to produce a disease dynamics more homogeneous from the spatial point of view (probably this should go to the discussion section).

c) Epidemic scenario three: Fully randomized Erdös-Rényi network

In the scenario 3 we investigate how a network, with the same number of nodes and approximately the same number of links as the observed flight network, but build instead by a process that produced theoretically a fully random connections between nodes, produces epidemic outbreaks. This way we can compare the disease dynamics that is produced by a theoretically complete random network with our reported network. Any deviation or difference of the spatio-temporal dynamics of the disease spread between the twos is an indication that the observed network was not "created" as a full random network –not at least by a process such as the Erdös-Rényi algorithm used here. Then, to build random networks keeping the constituent elements of the original observed network, i.e. both the number of nodes and links, so they would be comparable, we have used, as mentioned, the algorithm to assemble a random Erdös-Rényi $G(n,p)$ graph*** , see [12, 17] for details.

* The average (shortest) path length, together with the clustering coefficient and the degree distributions of the nodes are perhaps the most relevant tenets to characterize a complex network topology. For an exhaustive introduction see [17].

** As mentioned, the shortest average path length is a common measure for describing network topology and is defined as the average of the shortest paths between any two nodes in the network. Let $d(v_1, v_2)$ be the shortest distance or path between the vertices $v_1, v_2 \in V$, where V is the set of all vertices or links of the network. The average (shortest) path length is defines as:

$l_G = \frac{1}{n(n-1)} \sum_{i \neq j} d(v_i, v_j)$, where n is the number of links[35].

*** We use an algorithm to produce and uncorrelated random Erdös-Rényi $G(n,p)$ graph where n is the number of nodes and p is the probability of link formation so n nodes are connected through l edges which are chosen randomly from $n(n1)/2$ possible configurations. Every pair of nodes are connected with probability p. The total number of edges is a random variable with expected value $np(n1)/2$ hence $p = 2l/(n(n1))$. Thus introducing the observed number of nodes and observed number of links as n, l for the expression for p we obtain our probability of connection to generate a comparable fully random network to that of the observed network.

For each simulation we generate a $G(n,p)$ random graph by using the number of nodes n of the observed network and the corresponding link formation probability p that corresponds to the observed number of links and number of nodes. We then assign to the new network the nodes properties of local population sizes that correspond to the locations of the real network. We seed an infected individual at the Saint Martin node and simulate an epidemic. We save the generated artificial outbreak and proceed to repeat to generate a new random simulation. The Figure (5-6), right down plot, shows the epidemic curves of infected individuals of 5000 simulations (in light-red) of a fully randomized network based in both the number of links and nodes of the observed network model of flight connections. The number of links are kept the same as the observed network based on the flights data, but they are rearranged in a random manner according the Erdös-Rényi algorithm.

In comparison to the outbreak in observed network, the outbreaks simulated on the randomized network exhibit a reduced overall epidemic length. Only 5 simulations out of 5000 had longer durations than the observed network. If we hypothesize that the observed flight connection network was not generated as a fully random network, because the establishment of airline connections, by common sense, does not follow random rules but rather the decisions to make flight connections are far from arbitrary, then we would expect qualitatively differences between epidemics produced by completely random networks and the observed network based on the flights pattern. Accordingly, we expected fully random networks with exponential degree distributions, such as the ones generated by the Erdös-Renyi algorithm, to have an average shortest path lengths that is shorter than these path lengths found in natural networks with a more organized topology that bear a mix between structured and random link geometries, and skewed degree distributions [12, 17]. This would explain the significant shorter epidemic durations of the full random model as compared with the observed model as evident in the Figure (??). Shortest average paths of the fully random networks would be more efficient to diffuse the disease in the network yielding faster outbreaks with shorter extinction times and larger epidemic peaks, which is what we observe in the right-down plot of Figure (5-6).

d) Epidemic scenario four: start initial infection at each of the different network nodes

In this scenario the topology of the network model based on the flight connections is unmodified but each simulation is started at each of the nodes of the network as the initial infected location. The result is 30 different simulations that depict the scenario of the outbreak as if it were started at these locations, this is shown in the left-top plot of Figure (5-6). We notice that this scenario curves, when compared with the other scenarios, somehow retain both epidemic patterns observed in the outbreak produced by Saint Martin, namely the multi-peak curves of aggregated infecteds and the overall epidemic length.

Since the topology of the network is not modified and only the starting point of the infection process is (randomly) changed, the observation that the epidemic patterns show by the observed Saint Martin scenario are kept in the simulations indicate that the actual topology has an important role in the spatio-temporal spread of the CHIKV in the first stages of the outbreak in the Caribbean.

Intervention scenarios (in progress)

Scenario	Infecteds at Peak (mean/sd)	Time to Peak (days mean/sd)	Accumulated Infecteds at peak (mean/sd)
Scenario 1	577.2/104.1	129.6/30.2	363192.2/133459.7
Scenario 2	1032.4/0.0 *	80.0/0.0 *	338688.1/0.0 *
Scenario 3	734.8/166.6	103.1/18.6	327661.9/97152.9
Scenario 4	497.2/104.9	152.8/44.0	439021.1/129005.4

Discussion(in progress)

Methods

The model formulation

We have formulated a model for the initial outbreak of chikungunya infection in the populations of the Americas that started in the past December of 2013. The model is an epidemic compartmental model and is built as a set of coupled SEIR (i.e., susceptible-exposed-infected-recovered classes) ordinary differential equations (ODEs) linked by a host mobility network modeled after a dataset for commercial flights. Thus, the topology of the mobility network that interconnects the equations that describe the infection dynamics represents the flight connection network of the region at the time of the outbreak –this data was taken from the website openflights.org [22]. Therefore, this topology reflects an approximation to the real mobility pattern of people in the region during the epidemic event. The epidemic SEIR model is structured as a network where each node of the network represents a local population and is modeled as and homogeneous deterministic SEIR ODE which is in turn coupled with other similar SEIR nodes via migration rates given by the link strengths of the network of the flight connection.

The assumptions of the model are the followings:
(a) Each node is a homogeneous SEIR ODE system that represents the local epidemic dynamics of a particular closed population in an area (island, country, territory, etc.);
(b) Nodes are coupled with other nodes by migration rates of hosts;
(c) The network of flight connection is a good approximation for the pattern ofthe human host mobility among the locations in the Caribbean area;
(e) Mosquitoes are in excess and therefore their population dynamics does not matter for the infection dynamic in the host* . Under this assumption the infection term, i.e. the force of infection times the number of susceptible individuals λS, simply depends on the number of susceptible and infectious hosts, $\beta SI/N$, as there is always enough mosquitoes to infect people (β is the instantaneous transmission rate of infection, S, I, N are the number of susceptible, infectious and total individuals respectively); and
(f) Mosquito moving between locations is negligible.

A consequence of these assumptions is that human hosts are arguably the only carriers of the geographical spreading of CHIKV in the region. The mathematical description of the SEIR model is given by the following couple ODE system:

* This assumption renders the vector-borne model into a host-to-host infection process in the practice, but the difference would be in the transmission rate β which integrates mosquito mediated infection as a constant.

$$
\begin{aligned}
\frac{dS_i}{dt} &= -\beta_i \frac{S_i I_i}{N_i} + \sum_j \tau_{ij} S_j - \rho_i S_i, \\
\frac{dE_i}{dt} &= \beta_i \frac{S_i I_i}{N_i} - \sigma_i E_i + \sum_j \tau_{ij} E_j - \rho_i E_i, \\
\frac{dI_i}{dt} &= \sigma_i E_i - \gamma_i I_i + \sum_j \tau_{ji} I_j - \rho_i I_i, \\
\frac{dR_i}{dt} &= \gamma_i I_i + \sum_j \tau_{ji} R_j - \rho_i R_i, \\
i, j &= 1, \ldots, V.
\end{aligned}
\tag{1}
$$

where the indexes i, j denote the node locations and V is the total number of nodes (or vertices). We have then $4V$ couple equations. In our particular investigated case there are 30 nodes, so we set a model of 120 coupled equations. S_i, E_i, I_i, R_i denote the local population fractions of susceptible, exposed, infected, recovered, and $N_i = S_i + E_i + I_i + R_i$ is the total population of human hosts at each location i. We assumed in (e) of the previous model assumption list that the per capita transmission rate integrates the local mosquito dynamics.

These assumptions hypothesizes that the main mechanism for the spreading of chikungunya during the outbreak in the Caribbean is essentially a process driven by the human host mobility carrying the virus to remote locations and that a flight network can capture the spatio-temporal dynamics in the early stages of the outbreak. Additionally, the assumptions permit the setting of a relatively simple epidemic model with few infection parameters to fit the data -essentially the transmission rate of infection β, the rate of moving from exposed to infection class σ, and the recovery rate γ. The infection period can be computed as $1/\gamma$ and the latent period as $1/\sigma$ [14]. This simple setting with few parameters to estimate produces a convenient framework for policy makers and can serve as an operational theory to create strategic control in similar settings of geographical spreading of infections in loosely connected naive territories.

The data-based network model

The Figure 1 shows the investigated region. All the airports, together with their flight connections, reported in openflights.org [22] that lie within this area are used to build the nodes (or vertices) and links (or edges) of the network model. Note that the chikungunya incidence data reported and used in this work is aggregated by political regions such as countries or territories. Since there are typically several airports within each of these political region for which we have a single aggregated data set for the infection incidence, the corresponding airports located in the same political locations have been aggregated and their respective links summed accordingly. The result is a flight network model with the number of commercial airlines that connects the regions for which we have available chikungunya data.

The Figure (1, 4 top-right) shows the network model used for coupling each of the local node SEIR sub-models. Each node of the network models a location for which we have data on reported incidence of chikungunya (both confirmed and suspected cases).

For setting the model's initial conditions population sizes for each nodes has

been used corresponding with real population sizes of census information for locations these nodes represent. The values of population sizes of the nodes, that represent the local populations of the model, were taken from the census information of their respective island or countries found in Wikipedia [35]. The existence of a link between any two nodes indicates that there is at least one commercial airline operating between them for the time of the outbreak. Each link has associated a *strength*. The link strength between any two locations is directly proportional to the number of commercial airline operating between them (i.e. the number of companies flying) and is thought as to be directly proportional to the number of passengers traveling, i.e. this is integrated as the migration rate between two nodes in the mobility network. Each node has associated a population dynamics corresponding with its own population size and is governed by the equations of the model (1). The countries/islands represented as a local populations by the network nodes are the following: Puerto Rico, Dominican Republic, Jamaica, Haiti, Cuba, Cayman Islands, Bahamas, Colombia, Venezuela, Antigua and Barbuda, Barbados, Dominica, Martinique, Guadeloupe, Grenada, Virgin Islands, Saint Kitts and Nevis, Saint Lucia, Aruba, Bonaire, Curacao, Sint Eustatius, Sint Maarten, Anguilla, Trinidad and Tobago, British Virgin Islands, Saint Vincent and the Grenadines, Montserrat, Saba, and Saint Barthelemy. The list sums up 30 countries/regions.

Chikungunya incidence data

For validating the model (Equations set 1) we have used the chikungunya incidence case data as reported by the Project Tycho Data for Health (https://www.tycho.pitt.edu/) for the outbreak that started in the Caribbean in December 2013. The data set reports the incidence of both confirmed and suspected cases by country/region mostly on a weekly basis. The aggregated of all these reported incidence is plotted in the Fig. 2.

The fitted SEIR-network model

The parameters of the model (1) for the transmission rate, the rate of moving from exposed to infectious class (the inverse of the latent period), and the recovery rate (the inverse of the infectious period), β, σ, γ respectively, have been fitted to the observed data shown in figure (2). The method used is that of the least squares approach. Numerical solutions of (1) have been produced for different values of β, σ and γ from an initial parameter space build up out discrete intervals around initial value estimates of the parameters. The produced solutions -which are nearly continuous since they are numerical solutions of an ODE system-, have been then discretised and solutions points have been sampled to match both the number of points of the vector of observed data and their occurrences in time. In this way, the point-to-point distance between model solution vectors and the observed data vector can be computed, and both observations and model solutions be compared. A parameter space is then explored by computing solutions of the system (1) with combinations of parameter values of the space and the vector distances are compared to the vector of observed data. Intervals around the initial parameter estimates are established and then discretised, that is they are divided in ten points between the interval bounds. This produce a parameter space grid of 10^3 parameter combination points. Computations of the combinations of β, σ, γ on the parametric space grid are performed and the set of values that produces the solution with the shortest distance to the observation vector is selected and stored. This computation fulfills one iteration. A new iteration is then started by setting new intervals which are $1/10$th the size of the previous ones but centered around the new estimates selected and stored in the previous iteration. Again the new (smaller) intervals are divided in ten points which serves as a new parameter space of again 10^3 points for computing and selecting a new parameter set that again produces the shortest distance to the observation vector. This procedure is performed several times until convergence of the computed shortest distance

between a solution vector and the observation vector is observed. We have found that for the initial parameters estimates we use, the solution of the previously described iterative least squares fitting method converges after three iterations. That is, the fittest set of parameters β, σ, γ is selected out of $3 \times 10^3 = 3000$ points of the parameter space.

Finally, we have fitted two parameter sets by the least squares method previously described to the time series of the initial 203 days initial chikungunya observations (from 2013-10-20 to 2014-05-11). This time series represents the early stages of the outbreak. The initial values of the parameter set for fitting both the short and the long series is the same and is as follows:

$\beta = .81$, $\sigma = .78$, and $\gamma = .43$. The initial interval bounds are the initial parameter estimates plus/minus 0.5.

References

[1] Ling Bian. A conceptual framework for an individual-based spatially explicit epidemiological model. Environment and Planning B: Planning and Design, 31(3):381–395, 2004.

[2] Gianandrea Borgherini, Patrice Poubeau, Annie Jossaume, Arnaud Gouix, Liliane Cotte, Alain Michault, Claude Arvin-Berod, and Fabrice Paganin. Persistent arthralgia associated with chikungunya virus: a study of 88 adult patients on reunion island. Clinical Infectious Diseases, 47(4):469–475, 2008.

[3] Gianandrea Borgherini, Patrice Poubeau, Frederik Staikowsky, Manuella Lory, Nathalie Le Moullec, Jean Philippe Becquart, Catherine Wengling, Alain Michault, and Fabrice Paganin. Outbreak of chikungunya on reunion island: early clinical and laboratory features in 157 adult patients. Clinical Infectious Diseases, 44(11):1401–1407, 2007.

[4] SW Brighton, OW Prozesky, and AL De La Harpe. Chikungunya virus infection. a retrospective study of 107 cases. South African medical journal= Suid-Afrikaanse tydskrif vir geneeskunde, 63(9):313–315, 1983.

[5] CARPHA. Total reported cases of chikungunya in caribbean countries or territories and neighbouring mainland countries, 2014.

[6] CARPHA. Total reported cases of chikungunya in caribbean countries or territories and neighbouring mainland countries, 2016.

[7] Jean-Paul Chretien and Kenneth J Linthicum. Chikungunya in europe: what's next? The Lancet, 370(9602):1805–1806, 2007.

[8] Carlos J Dommar, Rachel Lowe, Marguerite Robinson, and Xavier Rod´ o. An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak. Acta tropica, 129:61–73, 2014.

[9] Richard Durrett and Simon Levin. The importance of being discrete (and spatial). Theoretical population biology, 46(3):363–394, 1994.

[10] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. Nature, 429(6988):180–184, 2004.

[11] Henryk Fuks and Anna T Lawniczak. Individual-based lattice model for spatial spread of epidemics. Discrete Dynamics in Nature and Society, 6(3):191–200, 2001.

[12] Matthew O Jackson. Social and economic networks. Princeton university press, 2010.

[13] PG Jupp and BM McIntosh. Chikungunya virus disease. The arboviruses: epidemiology and ecology, 2:137–57, 1988.

[14] Matt J Keeling and Pejman Rohani. Modeling infectious diseases in humans and animals. Princeton University Press, 2008.

[15] MJ Keeling and CA Gilligan. Bubonic plague: a metapopulation model of a zoonosis. Proceedings of the Royal Society of London B: Biological Sciences, 267(1458):2219–2230, 2000.

[16] Robert S Lanciotti, Olga L Kosoy, Janeen J Laven, Amanda J Panella, Jason O Velez, Amy J Lambert, and Grant L Campbell. Chikungunya virus in us travelers returning from india, 2006. Emerging infectious diseases, 13(5):764, 2007.

[17] Newman MEJ. Networks: an introduction. Oxford University Press, Oxford, 2010.

[18] Ruth M Myers and Donald E Carey. Concurrent isolation from patient of two arboviruses, chikungunya and dengue type 2. Science, 157(3794):1307–1308, 1967.

[19] Mark EJ Newman. Spread of epidemic disease on networks. Physical review E, 66(1):016128, 2002.

[20] Suchitra Nimmannitya, Scott B Halstead, Sanford N Cohen, Mark R Margiotta, et al. Dengue and chikungunya virus infection in man in thailand, 1962–64. i. observations on hospitalized patients with hemorrhagic fever. American journal of tropical medicine and hygiene, 18(6, Pt. 1):954–71, 1969.

[21] University of Pittsburgh. Project tycho data for health, 2015.

[22] Open-Flights. www.openflights.org, 2014.

[23] Ann M Powers, Aaron C Brault, Robert B Tesh, and Scott C Weaver. Re-emergence of chikungunya and o'nyong-nyong viruses: evidence for distinct geographical lineages and distant evolutionary relationships. Journal of General Virology, 81(2):471–479, 2000.

[24] Benjamin Queyriaux, Fabrice Simon, Marc Grandadam, R´ emy Michel, Hugues Tolou, and Jean-Paul Boutin. Clinical burden of chikungunya virus infection. The Lancet infectious diseases, 8(1):2–3, 2008.

[25] Jonathan M Read and Matt J Keeling. Disease evolution on networks: the role of contact structure. Proceedings of the Royal Society of London B: Biological Sciences, 270(1516):699–708, 2003.

[26] Steven Riley, Christophe Fraser, Christl A Donnelly, Azra C Ghani, Laith J Abu-Raddad, Anthony J Hedley, Gabriel M Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q Thach, et al. Transmission dynamics of the etiological agent of sars in hong kong: impact of public health interventions. Science,

300(5627):1961–1966, 2003.

[27] Marguerite Robinson, Anne Conan, Veasna Duong, Sowath Ly, Chantha Ngan, Philippe Buchy, Arnaud Tarantola, and Xavier Rodo. A model for a chikungunya outbreak in a rural cambodian setting: implications for disease control in uninfected areas. PLoS Negl Trop Dis, 8(9):e3120, 2014.

[28] Marion C Robinson. An epidemic of virus disease in southern province, tanganyika territory, in 1952–1953. Transactions of the Royal Society of Tropical Medicine and Hygiene, 49(1):28–32, 1955.

[29] Kibet Sergon, Charles Njuguna, Rosalia Kalani, Victor Ofula, Clayton Onyango, Limbaso S Konongoi, Sheryl Bedno, Heather Burke, Athman M Dumilla, Joseph Konde, et al. Seroprevalence of chikungunya virus (chikv) infection on lamu island, kenya, october 2004. The American journal of tropical medicine and hygiene, 78(2):333–337, 2008.

[30] Kibet Sergon, Ali Ahmed Yahaya, Jennifer Brown, Said A Bedja, Mohammed Mlindasse, Naphtali Agata, Yokouide Allaranger, Mamadou D Ball, Ann M Powers, Victor Ofula, et al. Seroprevalence of chikungunya virus infection on grande comore island, union of the comoros, 2005. The American journal of tropical medicine and hygiene, 76(6):1189–1193, 2007.

[31] Daouda Sissoko, Denis Malvy, Khaled Ezzedine, Philippe Renault, Frederic Moscetti, Martine Ledrans, and Vincent Pierre. Post-epidemic chikungunya disease on reunion island: course of rheumatic manifestations and associated factors over a 15-month period. PLoS Negl Trop Dis, 3(3):e389, 2009.

[32] J Erin Staples, Robert F Breiman, and Ann M Powers. Chikungunya fever: an epidemiological review of a re-emerging infectious disease. Clinical Infectious Diseases, 49(6):942–948, 2009.

[33] Konstantin A Tsetsarkin, Dana L Vanlandingham, Charles E McGee, and Stephen Higgs. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLOS pathog, 3(12):e201, 2007.

Figures

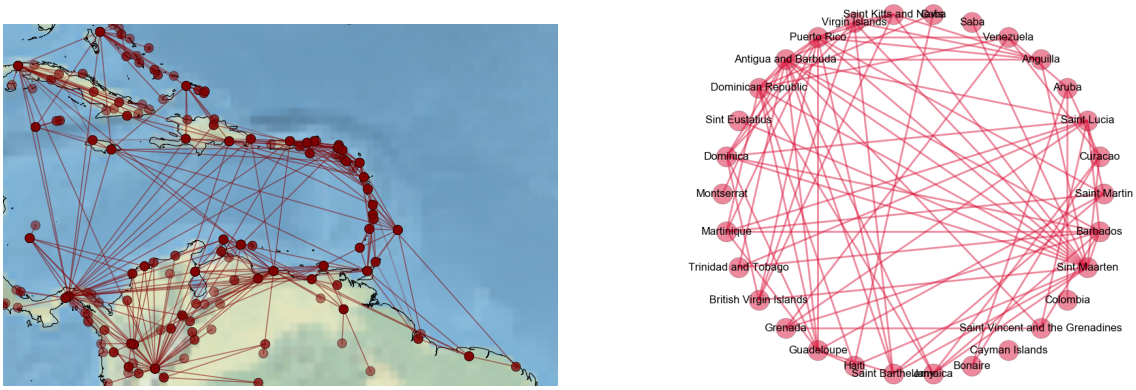

Figure 1: Modeled region and model network based on commercial airline traffic during the years 2013-2014 (www.openflights.com).

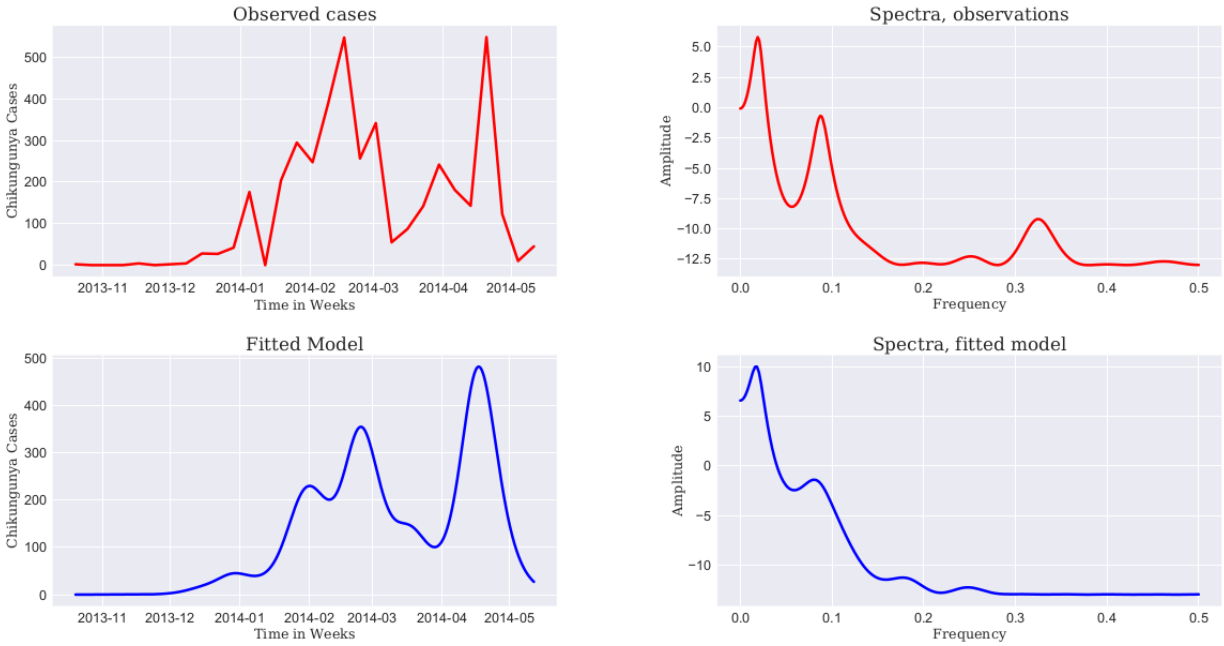

Figure 2: Aggregated reported chikungunya cases of the first 300 days of the outbreak and spectra of the observed time series (left up and bottom respectively); fitted (aggregated) solution of the network-SEIR model and spectra of the solution (right up and bottom respectively).

Figure 4: Epidemic wave. The profile of the solutions of the network-SEIR model is shown here for each individual nodes. The curves with larger amplitude correspond to simulated exposed individuals whereas the smaller amplitudes correspond to infected individuals.

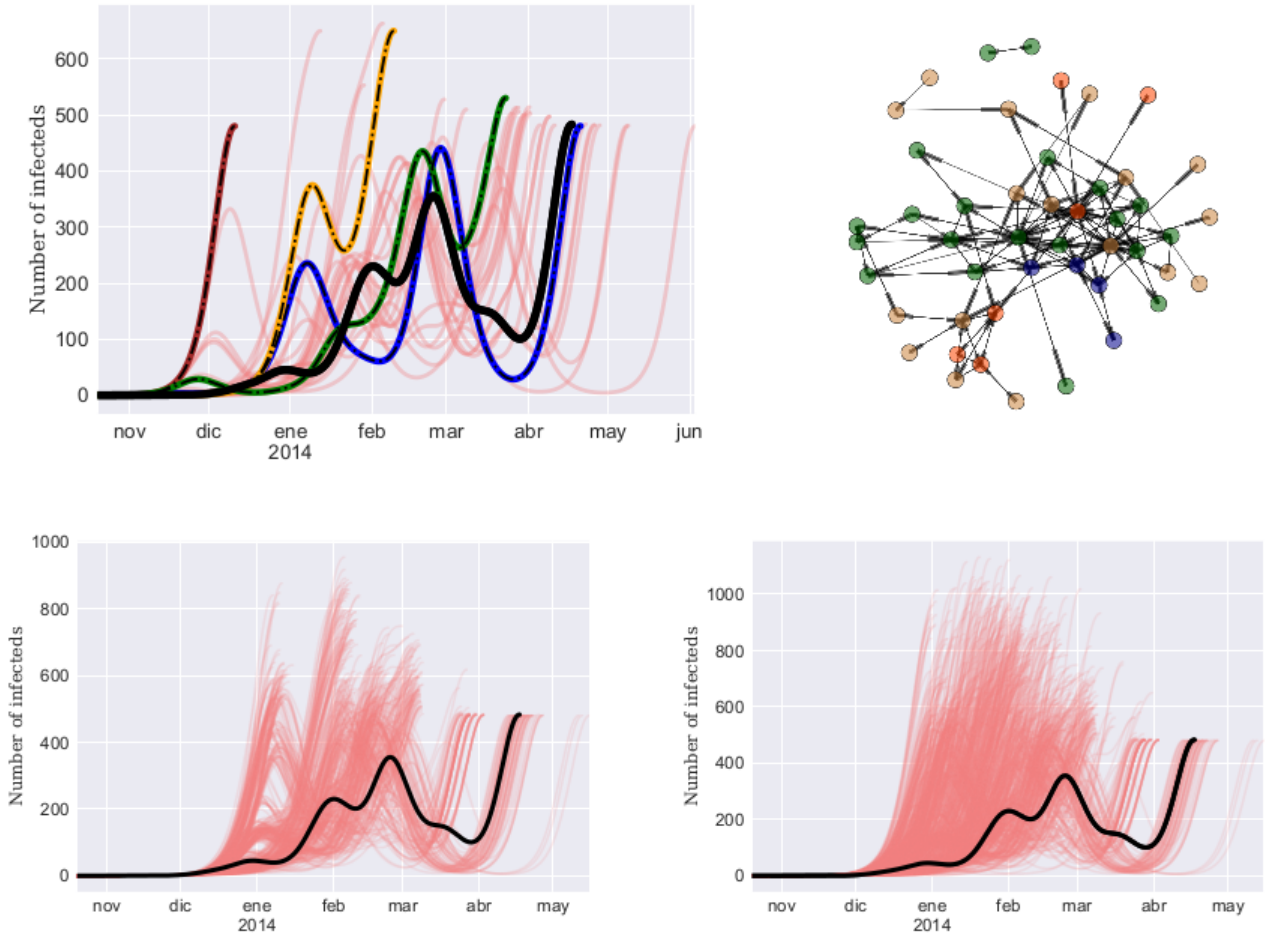

Figure 5: Simulation scenarios. Top left is the network model of airline connections and colors classify nodes regard official lnaguages spoken at each region: blue (French), orange (Ducth), latin-spanish(brown), and anglo/english (green). Top right are simulations (light red curves) of the fitted model where infection was initiated at each of the nodes of the network model other than in Saint Martin (which is represented by the solid black curve), the colored curves correspond to the outbreaks which start at the largest node population-wise of the nodes grouped by spoken language, which corresponding colors. Bottom left represent the scenario where the links for the initial Saint Martin have been randomized (light red); and bottom left represnt the scenario where the links for all nodes have been randomized following a Erdös-Renyi random network model (ligh red). The modeled outbreak initiaed in Saint Martin is the black solid curve.

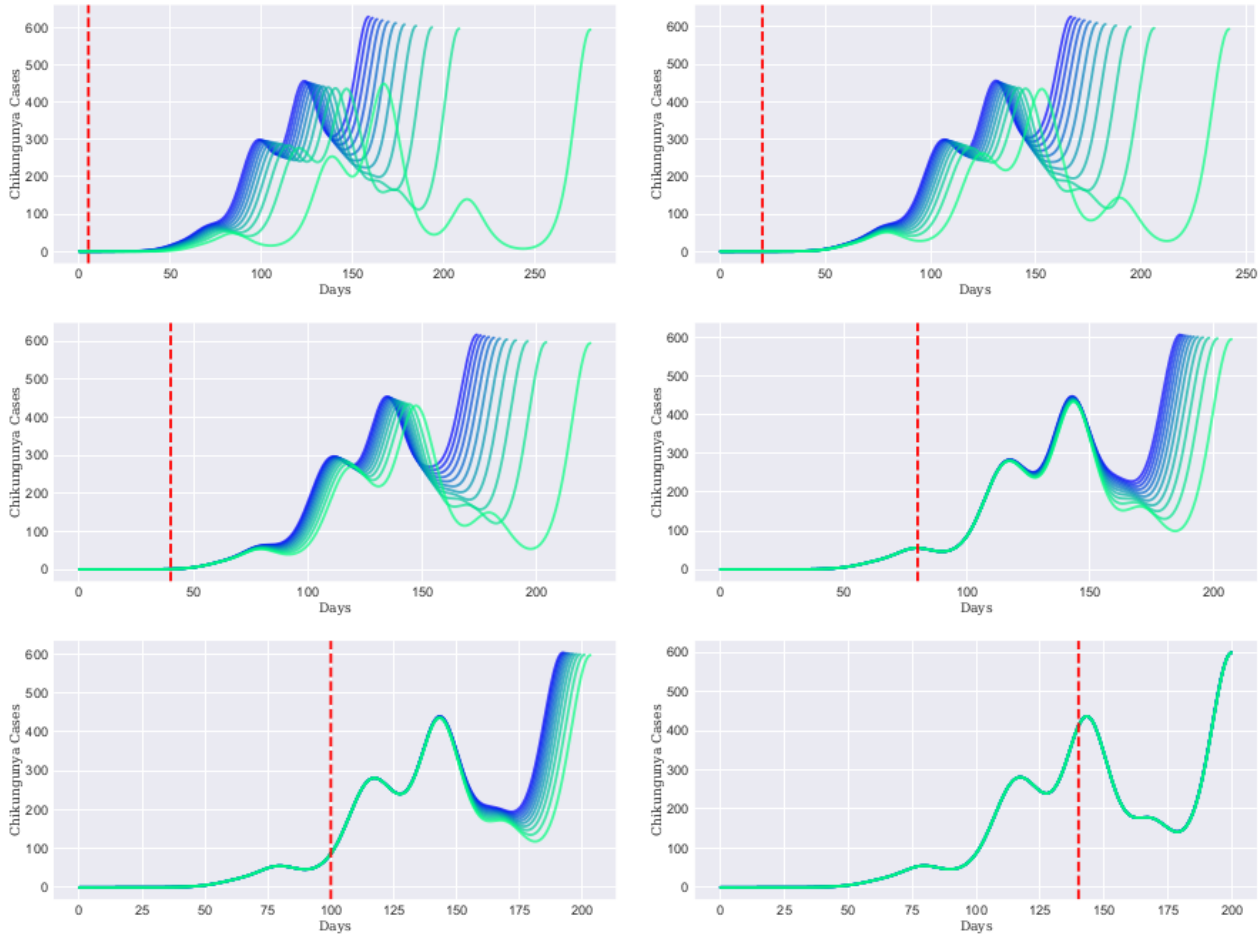

Figure 7: . Intervention scenarios. Level of reductions of air traffic have been performed on the fitted network model to assess epidemic severity under control scenarios. Ten reduction levels were performed: from 100% (total cut off air traffic), 90%, ..., up to 0 % (not cut off at all). Simulations of the network model were executed with these interventions and outbreak curves were plotted. The more blue indicates higher level of intervention (higher cut off of the air traffic). The vertical red line indicate the time when the interventions were made.

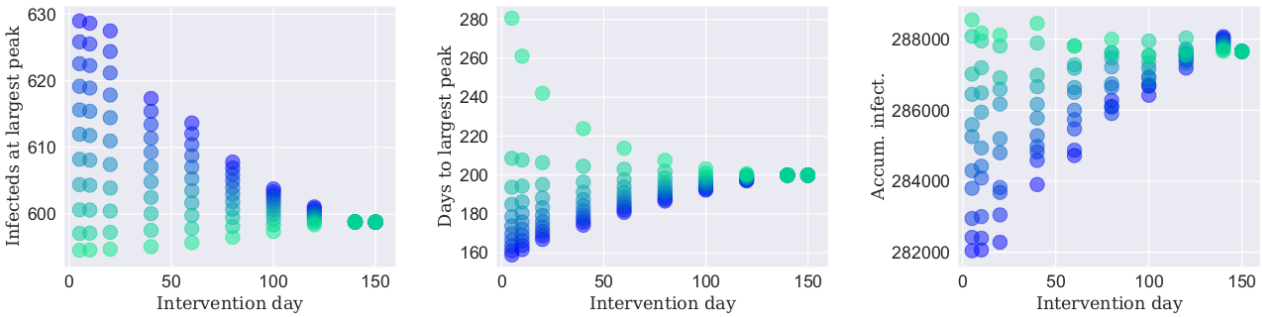

Figure 8: Intervention scenarios. Right plot indicates the epidemic severity as a function of both the day of intervention since the epidemic day zero (in the x-axis) and the level of intervention (indicate by the intensity of blue- the more blue the largest the intervention the more green the less intervention). The plot of the right measures the epidemic severity as the number of infected individuals at the maximum epidemic peak, the plot in the center measures the epidemic severity as the number of days to the maximum epidemic peak, and the plot of the left measures epidemic severity as the number of accumulated infected individuals at the maximum epidemic peak.