

# Scientific Reports: A commercial airline network model for chikungunya spread in the Caribbean

Carlos J. Dommar and Xavier Rodó

June 13, 2017

## Abstract

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 that started in the end of 2013 in the Caribbean region. The 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 have been investigated for such purpose. And finally simulations of the fitted model of a simple intervention measurement demonstrate the appearance of a limit time horizon where any level control efforts later than it will 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.

Example Abstract. Abstract must be under 200 words and not include subheadings or citations. (This is for Scientific Reports).

## 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

forest-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 propagates 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].

In this study we briefly describe and simulate the first 6 months of the expansion of CHIKV in the Caribbean region to try to understand what the main factors are in determining how the propagation initially takes place. The temporal restriction seeks to trace the disease at the times when both primary and secondary infections were 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 determined by the structure of the population, herd immunity lo-

cally 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.

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.

## 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](http://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 sim-

ilar 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 of the 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<sup>1</sup>. Under this assumption the infection term, i.e. the force of infection times the number of susceptible individuals  $\lambda S$ , simply depends on the number of susceptible and infectious hosts,  $\beta SI/N$ , as there is always enough mosquitoes to infect people ( $\beta$  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:

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

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

---

<sup>1</sup>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  $\beta$  which integrates mosquito-mediated infection as a constant.

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  $\beta$ , the rate of moving from exposed to infection class  $\sigma$ , and the recovery rate  $\gamma$ . The infection period can be computed as  $1/\lambda$  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

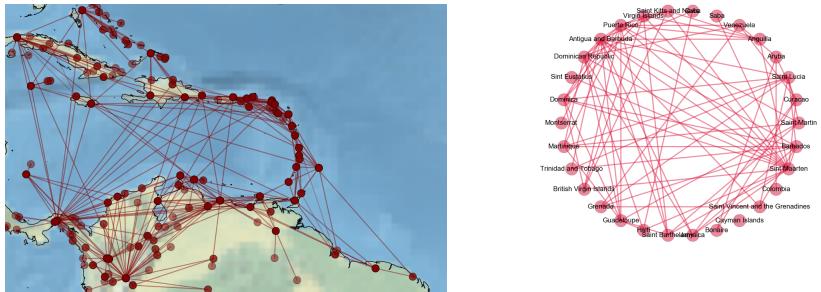


Figure 1: Global map showing the Caribbean area investigated (upper map). All locations and airports lying in the squared area are implemented in the SEIR network model. We show all the airport reported by openflights.org with their respective flight connections represented as links (bottom figure).

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 ?? shows the network model used for coupling each of the local-node SEIR submodels. Each node of the network models a location for which we have data on reported incidence of chikungunya (both confirmed and suspected cases). 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 (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.

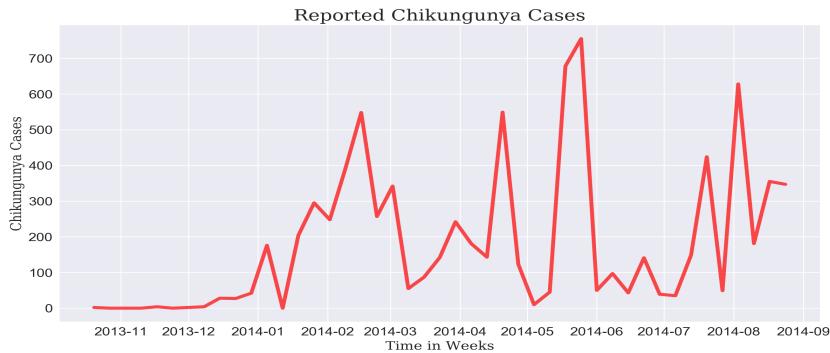


Figure 2: The incidence data. The plot shows the total aggregated number of cases (both confirmed and suspected) for the chikungunya outbreak in the Caribbean for the initial 300 days period (from October 20th 2013 to August 24th 2014) according the database of the Project Tycho (<https://www.tycho.pitt.edu/>).

### 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),  $-\beta, \sigma, \gamma$  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  $\beta, \sigma$ , and  $\gamma$  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  $\beta, \sigma, \gamma$  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 1/10th 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  $\beta, \sigma, \gamma$  is selected out of roughly  $3 \times 10^3 = 3000$  points of the parameter space.

Finally, we have fitted two parameter sets by the least squares method previously described to two observed time series: one time series for the very initial chikungunya observations (from 2013-10-20 to 2014-05-11), and other for the whole observed chikungunya incidence (2013-10-20 to 2014-08-24). Both time series represents the early stages of the outbreak, only that the first series accounts for the first 203 days, and the second series accounts for the first 308 days. Here we refer to them as the ‘short’ and the ‘long’ time series of incidence observation. The initial values of the parameter set for fitting both the short and the long series is the same and is as follows:  $\beta = 0.81$ ,  $\sigma = 0.78$  and  $\gamma = 0.43$ . The initial interval bounds are the initial parameter estimates plus/minus 0.5.

The Figure (??) shows the fitted model after three iterations for the observed series of the initial 308 days of the outbreak and the Figure (??) shows the fitted model for the initial 203 days. As previously mentioned, each of these fitting involved the computation of solutions of model (1) for 1000 points (

## Model results

As described in the previous 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 months of the outbreak in the Caribbean that started towards

the end of 2013 in Saint Martin Island.

The Fig 3 shows two fitted numerical solution (green curves) of (1): one model solution fitted for the first 203 days (in the middle panel), and one model solution fitted to the first 308 days of the outbreak (bottom panel). The real aggregated observed data for that time periods is shown in the upper panel (red plot). 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].

As shown in the picture (3) the model captures the epidemic ridged curve shape exhibited by the observed time series of aggregated observations. The multi-peaked pattern of the aggregated observed incidence is made of various local single epidemic curved peaks occurring at the several islands locations of the region that when are aggregated they render the time series shape as ridged and multi spiky (Fig 2). This can be observed when plotting together the individual observed time series for locations (not shown here). A plot with a representation of the modeled individual epidemic peaks at the level of node can be seen in the Figure(4). The fact that all these several peaks at different nodes (and geographical locations) that are occurring at different times are all coming for a single infected individual in particular space and time point is what let us recognize the process as a spatio-temporal epidemic wave.

The model capture the same ridged pattern in the node aggregated numerical solution by being able to simulate the same pattern shown in the observed data. Specifically, the model produces a distinct spatio-temporal outbreak with single peaked simulated epidemic curves at each discrete population modeled as a node that when plotted together produce the recognizable jagged curve seen in the real data (3). When plotted together the individual modeled epidemic curves of the nodes (not shown here) it becomes apparent that the multi-peaked shape of the aggregated model solution emerges from single peaked solutions on the several spatially apart nodes at different times. Just the same pattern as it is displayed in the observations. Therefore we may say that the model successfully captures the spatio-temporal mechanism observed in the data.

The table 1 shows the parameter values fitted for the short and long time series of observations used in the model 1. The epidemic wave spread has been simulated by solving numerically the SEIR network model (1) with the parameters estimated as described previously. 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. To simulate the observed scenario, a single infectious individual has been set at the node of Saint Martin at the epidemic day 0. These initial conditions are in accordance with the observed epidemic plot of the Chikungunya outbreak in the Americas in 2013.

The simulated geographical spread wave produced by the fitted model has been inspected. As mentioned, the figures (??, ??, 3) show the model simulations of the aggregated cases for all nodes/locations. The individual simulated epidemic peaks for each nodes/locations are shown in Figure (4). These graphics reveal

Parameter	Definition	Fitted Value
$\beta$	Transmission rate	$1.21* / 0.448* \text{ day}^{-1}$
$\sigma$	Incubation period	$0.354* / 0.665* \text{ day}^{-1}$
$\gamma$	Recovery rate	$0.568* / 0.269* \text{ day}^{-1}$

Table 1: Parameters of the epidemic model estimated by least squares from the observed incidence series. The values marked with (\*) corresponds to the estimates for the initial 203 days of the outbreak (see Figure ??), the values marked with (\*) corresponds to the estimates for the initial 308 days of the outbreak(see Figure ??).

the spatio-temporal structure of the epidemic wave simulated by the model by portraying the timing of the local highest epidemic peaks (for both the exposed and infectious classes). However, note that the relative magnitudes of the peaks appear similar. That is, the peaks show comparatively the same amplitude when they are not in reality; this is plotted this way for the sake of clarity. Albeit in the Figure (4) the number of cases cannot be comparable among the different nodes/locations plots, the axis for time is identical for all of them making the timing of the local epidemic onsets and epidemic peaks comparable indeed. Thus, as earlier mentioned, the relative time shifts of the epidemic peaks displayed by plotting the individual solutions of the nodes of the SEIR network model unveil the spatio-temporal structure of the observed epidemic wave captured by the model.

We then use the fitted SEIR network model 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.

We hypothesize that in an epidemic setting such as the 2013 chikungunya outbreak in the Caribbean, the geometry of the mobility pattern of the host individuals plays a definite role in the geographical spread of the disease as they 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 are naive 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. In the next sections we describe the scenarios.

### Epidemic scenarios

Four scenarios have been 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 connec-**

### tions of the initial infected node

In this scenario we have performed simulations of the fitted network SEIR model (1) where the connections of the initial infected node (Saint Martin) has been randomized. That is, the original links that represent flight connection of Saint Martin have been randomly changed along with its corresponding strengths, the number of links of the original observed network (6) 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 flights. The Fig. (??) shows 500 random simulations in which, at each single simulation, the original six connections of Saint Martin has been randomly changed. The red curve corresponds to evolution of number of aggregated infected individuals of the original network model based on the observed flight data. The curves corresponds to aggregate simulated cases of all the nodes/locations.

Visual inspection shows 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 blue epidemic curves in the Figure (??). The infection can go over many several paths producing a variety of epidemic geographic diffusion because of a seemingly small disturbance in the network structure. 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 calculate the number of possible epidemic waves as the combinatoric number  $\binom{29}{6} = 29!/29!(29 - 6)!$  that is, there are 475020 possible combinations and therefore so different epidemic waves. Due to computational limitations, we have randomly chosen and simulated 500 of this possible trajectories in the hope that this is a representative sample of all possible combinations.

We can notice that the epidemic duration of the outbreak produced by original model based on the observed flight connections lies in the tail (towards the longer duration) of the distribution of epidemic durations resulting out 500 random simulations. This apparently implies 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 Saint Martin. In other words, the epidemic dynamics can be sensitive to small changes and perturbations in the topology of the network mobility. In the plots of the Figure (??) statistical descriptions of the experiment of the epidemic scenario one are shown. We see that the number of infected individuals at the epidemic peak is mostly around 500 and then the number of infected and the peaks decays at what seems an exponential manner. The histogram does not show epidemic peaks with less than 500 infecteds however. The histograms for the time to reach epidemic peaks is not very informative though; it seems that most of the simulations peaked between 100 and 140 days after initial infection but a substantial number peaked at as earlier as

80 days and at as later as 180-200 days.

We have then explored a quantity that can be of interest in public epidemiology; we called the *population at risk on first degree*. We define the population at risk on first degree as the size of the total population that is **directly** connected with the initial population were the infection is first reported, with this initial population included. In our context, the population risk a first degree is the sum of all populations connected (by commercial flights) to the initial infected population, including the latter. This quantity measures the total population put in peril immediately after a reported case is confirmed. The simulations for epidemic scenario one produced a histogram (Fig. ?? left-middle panel) that reports that most of the population at risk on first degree has a size between  $\exp(10)$  to  $\exp(11)$ , i.e.  $2.2 \times 10^4$  to  $5.99 \times 10^4$  individuals. The total population of the studied geographical region is 121,913,502 individuals, therefore the model predicts a population at risk on first degree between 0.018% to 0.05% of the population of the region. These are the average population percentages that would be at risk of immediate contagion right after the first case is reported.

We also report three 2D histograms by contrasting pairwise the previously described ones. We report then: (a) maximum epidemic peak size vs. time to reach the peak; (b) maximum epidemic peak size vs. population at risk on first degree; and (c) time to epidemic peak vs. population at risk of first degree (Fig. ??). The 2D histogram (a) reports that the larger epidemic peak sizes occurred mostly between 150 to 160 days after the start of the outbreak and many peaks of 500-650 individual sizes occurred between 100 and 10 days after the start of the outbreak. 2D histogram (b) reports that the largest sizes of population at risk on first degree, between  $\exp(9)$  to  $\exp(11)$ , occurred when the sizes of the largest epidemic peaks were at around 500-550 individuals. Finally the 2D histogram (c) reports that the highest levels of population at risk on first degrees,  $\exp(10)$  to  $\exp(11)$ , happen when the time to reach the highest epidemic peaks is around 80 days. Population at risk of around  $\exp(9)$  to  $\exp(10)$  seems to be distributed more or less evenly across epidemic peaking times between 100 to 180 days.

### 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. Figure ?? shows the epidemic curve of 500 simulations of this scenario. As one would expect, the epidemic curve of aggregated individuals exhibits a typical pattern of an outbreak of an homogeneous non-spatial model. Creating a connection of the initial infected node to all other nodes in the network produce an infection pattern close to that of an homogeneous unstructured SEIR model. This is unmistakable apparent in the plot (??) since the epidemic curve lacks the jagged structure of an explicitly spatial wave shown in the solution of the fitted models (??,??).

The Figure (??) plots 500 simulations (in red) which appears as a single curve. The variation due to the randomization of the connection strengths is virtually non existent as seen in the plot. Since, in this scenario, what we (randomly) vary at each single simulation are the link strengths while keeping the topology constant, 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 populations. That is, that topology matters more than dynamics. Therefore, 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 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<sup>2</sup> 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<sup>3</sup>. 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 scenario  $l_G^{scen2} = 1.754$ , which indicates that effectively the average shortest path has been decreases by connecting Saint Martin to all other nodes. This makes the contact structure of the transmission kernel to produce a disease dynamics more homogeneous from the spatial point of view.

Because, as discussed earlier, trajectories of the simulations of this particular scenario lie nearly on the same curve, practically no variation is then produced which render statistics futile in providing further information on the system, hence we do not take on the task to report any statistics as we do for the other cases.

### c) Epidemic scenario three: Fully randomized Erdős-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 designed as a full random network –not at least by

---

<sup>2</sup>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].

<sup>3</sup>As said, 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  $n$  s the number of links[35].

a process such as the Erdős-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 Erdős-Rényi  $G(n, p)$  graph<sup>4</sup>, see [12, 17] for details.

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 (??) shows the of the curves of infected individuals of 500 simulations (in red) of a fully randomized network based in the network model of observed 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 Erdős-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 500 had longer durations than the observed network. If we conjecture 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 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 to have an average shortest path lengths 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 plots of (??). Additionally, the histograms of Figure (??) reveal that the epidemic sizes are in general larger than any of the epidemic scenarios investigate here. The most common epidemic peak sizes observed in the Erdős-Rényi model are about 800-1000 individuals (typical peaks of scenario 1 is around 500 individuals), and the more frequent times to peak oscillates from 80 to 120 days (which are significant shorted than the values of 100 to 140 days for scenario 1). The typical population sizes at risk

---

<sup>4</sup> We use an algorithm to produce and uncorrelated random Erdős-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  $n(n - 1)/2$  possible configurations. Every pair of nodes are connected with probability  $p$ . The total number of edges is a random variable with expected value  $np(n - 1)/2$  hence  $p = 2l/(n(n - 1))$ . 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.

on first degree lies between  $\exp(10)$  to  $\exp(12)$  (which is significant larger than the values  $\exp(10)$  to  $\exp(11)$  for scenario 1). According the 2D histograms we can see that the largest epidemic peaks (900 to 1100 individuals) occurs when the time to reach these peaks are about 90 to 110 days from the start of the epidemic. The highest values for population size at risk on first degree (around  $\exp(11)$  and more) occurs when the maximum epidemic peak is around 800-900 individuals. And finally, these highest values of population at risk on fist degree coincide when the times to epidemic peaks are at 60 days from the start of the infection.

#### d) Epidemic scenario four: Initial infected node location chosen randomly.

In this scenario the topology of the network model based on the flight connections is unmodified but the initial infected location or node is chosen randomly for each simulation realization. The result of a simulation of 250 random realizations is show in Fig. ???. We can notice that both epidemic patterns observed in the outbreak produced by starting in the node corresponding with the real observation Saint Martin, i.e. multi-peak curves of aggregated infecteds and overall epidemic length, are still reproduced in this scenario.

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 are retained in the simulations indicate that the actual topology has an important role in the spatio-temporal deployment of the CHIKV in the first stages of the outbreak in the Caribbean. We then can consider this model to investigate some intervention scenarios to learn about the conditions that might help to mitigate similar outbreaks in future situations.

Without specifying width for last column:

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

### Intervention scenarios

The calibrated and parameterised model can be use to explore simple cases of interventions in order to mimic the consequences of the decisions taken and public policies and strategies triggered in face of an icoming epidemic event. The plots shown here represent the effect of intervention actions on the SEIR network model of chikungunya for the Caribbean that started in November 2013 based in a set of coupled ODEs by means of a network topology that resembles the real traffic of commercial airline flights.

The intervention are based on evident actions that decision makers controlling air traffic can take, that is reducing flights numbers or flights activity at

some point in time after the outbreak has been recognized. This is modelled here as a reduction in the percentage or level of the connection strength of the air traffic network model at different times after day one of the outbreak, i.e. when a first infected is accounted.

Eleven levels –from 0%, 10%, 20%, …, %100– of “intervention actions” have been simulated. For example 0% means no intervention and the interaction strength is the same as the fitted model, 10% means that there is a 10% of reduction of the link strengths in the network, and so on.

Sixteen times to implement an intervention action after outbreak day one has been simulated to investigate the effect of the response quickness of the intervention on the overall outbreak. The plots represent the time evolution of the aggregated infected in all the nodes/islands/populations of the Caribbean network.

## Discussion

The Discussion should be succinct and must not contain subheadings.

## 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 Dommars, Rachel Lowe, Marguerite Robinson, and Xavier Rodó. 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 Mar-giotta, 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](http://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, Rémy 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, Fredéric 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.

- [34] Marie Vazeille, Sara Moutailler, Daniel Coudrier, Claudine Rousseaux, Huot Khun, Michel Huerre, Julien Thiria, Jean-Sébastien Dehecq, Didier Fontenille, Isabelle Schuffenecker, et al. Two chikungunya isolates from the outbreak of la reunion (indian ocean) exhibit different patterns of infection in the mosquito, *aedes albopictus*. *PloS one*, 2(11):e1168, 2007.
- [35] Wikipedia, 2017. [Online; accessed 10-January-2017].

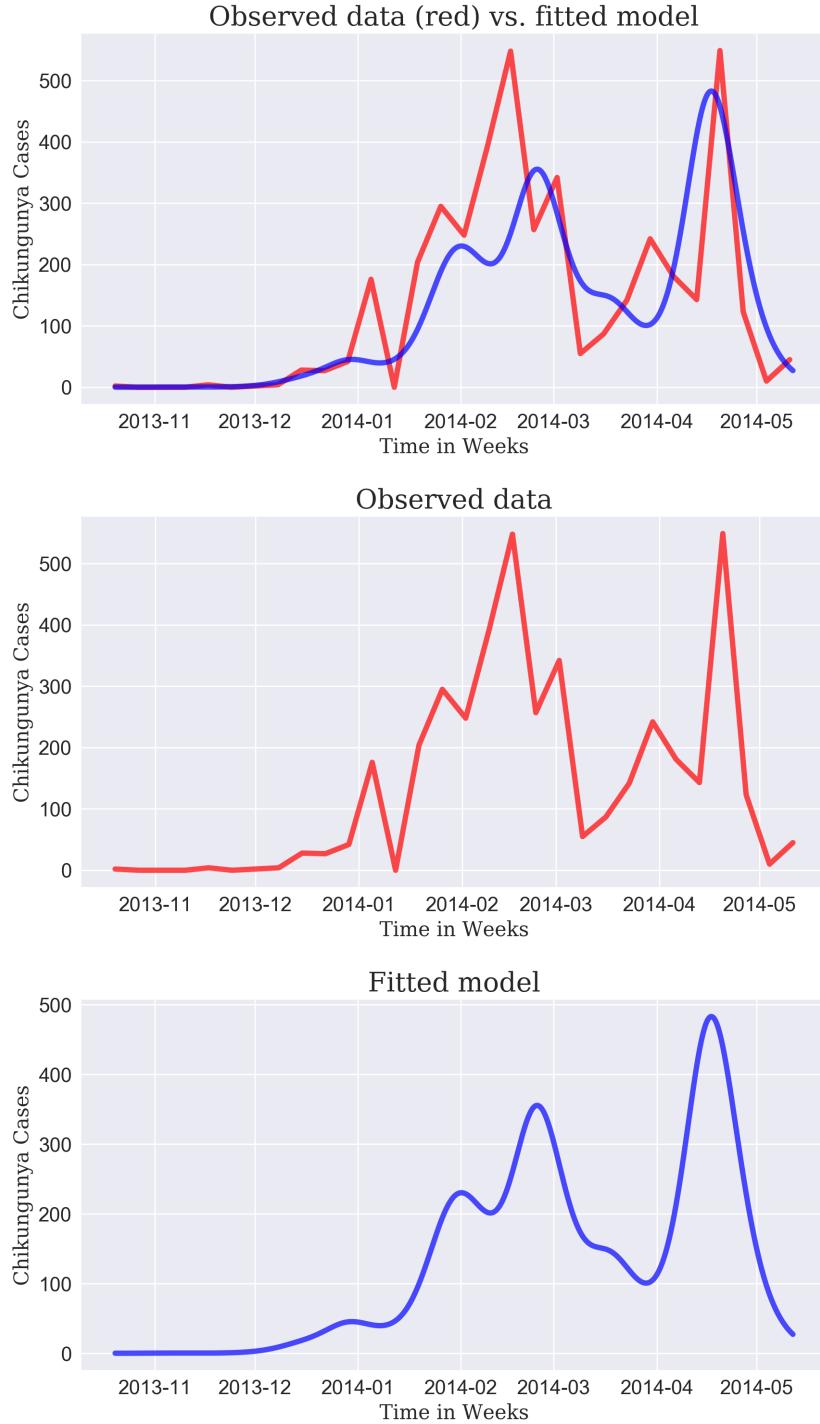


Figure 3: Fitted SEIR Mobility Network model compared to observed data. The numerical solution of the model captures the initial spreading wave of the observed aggregated incidence. The red time series shows the reported aggregated incidence data (confirmed and suspected cases) of chikungunya in the Americas during the first 308 days after first local transmission has been reported in Saint Martin Island in October 20th of 2013. The blue curve shows the SIR Network model numerical solution with the fitted parameters for the observed time series. Fitted parameters are  $\beta = 1.21$ ,  $\sigma = 0.354$ , and  $\gamma = 0.568$ .

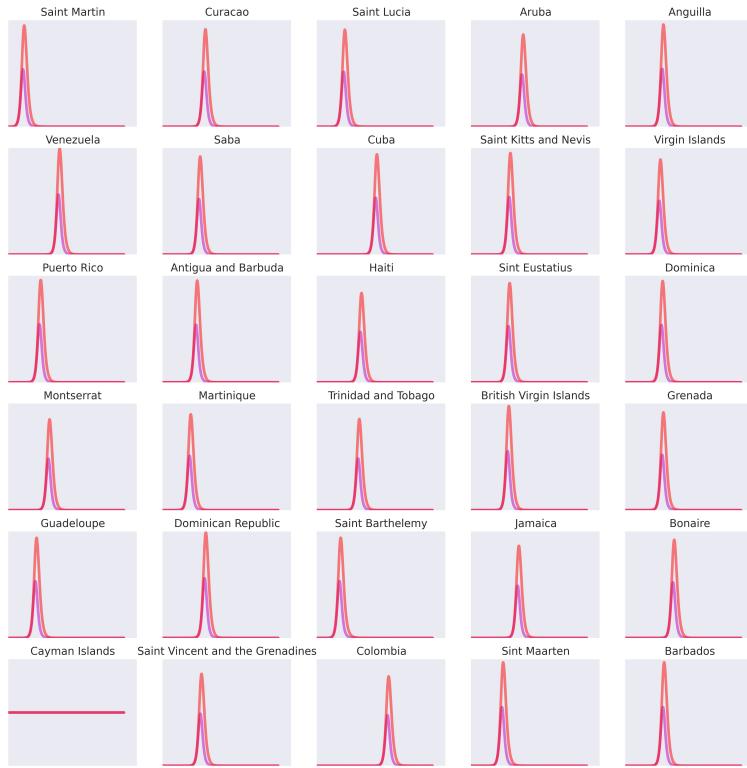


Figure 4: The epidemic wave. This figure shows the several simulated local epidemic curves of infected and exposed individuals (red and purple curves respectively) at each of the nodes that represent different locations in the Caribbean. At each one of the small plot-boxes the vertical axis represents individual cases and the horizontal axis is time. Note that amplitude scales are not relative and therefore the epidemic magnitudes cannot be compared, however their timings can indeed be. The epidemic wave becomes apparent as the local epidemics peak occurs at different times. The flat curve of Cayman Islands shows no outbreak there and this is in accordance with the case that the island is not connected to the rest of the set of islands on the space domain corresponding to the selected geographical region. Therefore the Cayman Islands are not hit by the infection in our simulations. This is consistent with the observation that no chikungunya cases are reported in the Cayman Islands during the simulated time period.

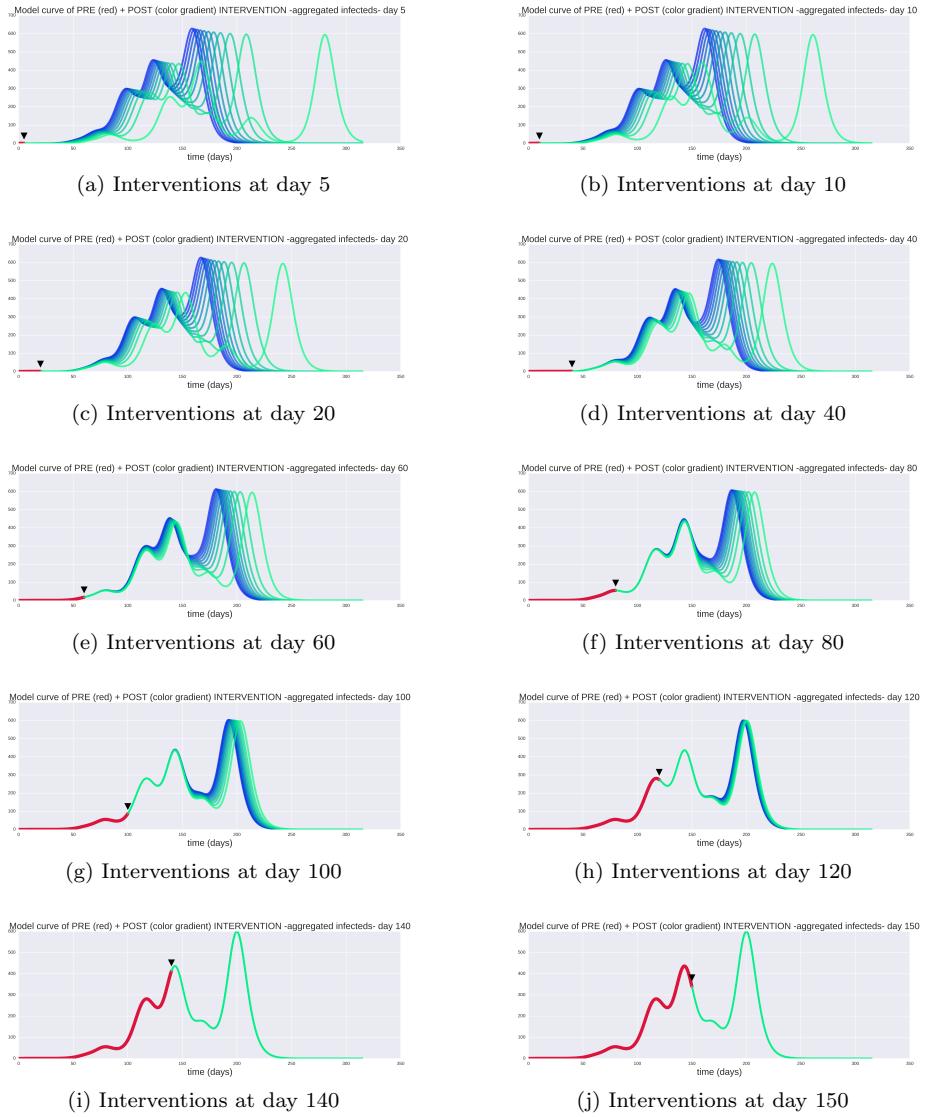


Figure 5: The effect of several levels of interventions at different times of the epidemic. The outcome of the SEIR network model fitted for the short data time series (??) is inspected under a simple intervention action such as mobility restriction by halting the flight traffic. We have simulated this intervention by decreasing the original migration rates at 10 different degrees of their original values; i.e. from 100% of the original migration rates which is the full original migration rate, going through 90%, 80%, ... all the way down to 0% that corresponds to no migration and isolation of the nodes. Remind that migration rates are modeled as the link strengths in the network model. Therefore in the scenario of 0% of the migration rates, we have a completely disconnected network depicting thus the maximum degree of intervention. The red curves plot the modeled chikungunya case incidence prior-intervention. The small black arrow points the exact time where the intervention has been done. The action of intervening by cutting down the migration rates is then held. The consequences of the different degrees on interventions are depicted by the color gradation of the epidemics curves: the curves colored towards the green have undergone less intervention (that is migration rate less decreased) with the greener curve having no intervention at all. On the other hand, the curves with colors turning towards the blue have undergone more intervention, and the bluest of them having complete intervention and rendering the network into a set of completely isolated nodes.