Understanding the role of sexual transmission in the spread of ZIKA virus using an individual-based interconnected population model

Tanvir Ferdousi[1,*], Lee Cohnstaedt[2], D McVey[2], and Caterina Scoglio[1]

[1]Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506, USA
[2]Center for Grain and Animal Health Research, USDA, Manhattan, KS 66502, USA
[*]tanvirf@ksu.edu

ABSTRACT

Zika virus has affected the world as a long-term threat. Modeling its transmission is important in order to facilitate forecasts and control measures. We propose a novel node-based interconnected population model to simulate both vectored and sexual transmission of Zika virus. Using a sexual contact network, we incorporate heterogeneous mixing in the host population with stochastic transmission for realistic predictions. We also incorporate climatic variations in our model, which affect the mosquito vector population and consequently the arbovirus transmission. We perform extensive simulations to understand the effects of sexual transmission rate and network topology on the spreading of infections. Sexual transmission contributes to the epidemic spread and under certain conditions, can sustain it up to several months without vectors. This can potentially lead to recurrences once the mosquitoes overwinter. We also find that sexual transmission can have a stronger effect when vectored transmission is relatively weaker due to climatic conditions. Our results show that vectored and sexual transmission affect the disease dynamics differently.

Introduction

Zika virus (ZIKV) is a positive-sense single-stranded RNA virus of the Flaviviridae family[1]. It is related to other flaviviruses such as Dengue, Yellow fever, West Nile and Japanese Encephalitis. ZIKV infection symptoms include acute fever, maculopapular rash, arthralgia and conjunctivitis. However, only 20% of the cases show symptoms[2], which poses a major problem in the timing of response measures. ZIKA infection in pregnant women can result into congenital microcephaly; a severe birth defect that causes underdeveloped brain in the newborns[3]. The transmission occurs primarily by infected mosquito vectors such as Aedes aegypti and Aedes albopictus. However, it can also spread via semen and blood[4].

In 2015, an outbreak was confirmed in Brazil. The infection quickly spread into many South American territories by the end of that year. The WHO declared Zika outbreak as a Public Health Emergency of International Concern (PHEIC) in February 2016. The first autochthonous case of Zika in the mainland US was reported in July 2016, in Miami, Florida. As of February 21st 2018, 5658 cases have been reported in the US States. In 2016, 228 autochthonous cases have been reported where the state of Florida itself is responsible for 97% of those cases, and the remaining 3% cases occurred in Texas[5]. There is very little knowledge about the spread of ZIKV in the Americas prior to WHO's declaration. Many parameters regarding the epidemic are still unknown. Recent studies assume those parameters using the estimates from Dengue outbreak data[6789].

As the disease is predominantly vector-borne, the outbreaks are strongly dependent on *Aedes* mosquito vector density and the climatic conditions[6]. The mosquito lifespan and extrinsic incubation period is dependent on temperature[8]. For these reasons, outbreaks occur only in those places where the vectors are abundant. Even in those locations, ZIKV outbreaks show strong seasonal variation due to the fluctuations in temperature and precipitations[7]. Not all people are equally susceptible to the infection. The availability of air-conditioning, mosquito repellents, situational awareness etc. affect the exposure to arboviruses. Socio-economic condition introduces another dimension of heterogeneity in the susceptible population[6]. ZIKV epidemics have slow growth rates[6]; however, it is still difficult to control due to low reporting rates and relatively slow screening process[10]. International travel has contributed to the spreading across continents[5].

There have been several attempts in modeling the outbreak of ZIKV. Kucharski et al. used the classical compartmental model to simulate the 2013-14 outbreak in French Polynesia[11]. The model assumed spatial homogeneity and no effect of seasonality. Castro et al. performed a risk assessment for Zika in Texas[7]. They estimated county level importation risks from socioeconomic and demographic data and used those to estimate ZIKV imports and the subsequent probabilities of triggering large outbreaks. Their work incorporated a multi-compartment model for the host alone (humans) in a stochastic branching

process. They fitted the Exposed class to consider the effects of vector-borne transmission[7]. Perkins et al. estimated epidemic attack rates for highly spatially resolved data[8] using the classical formulation of epidemic burnout[12]. This work projected a revised upper limit of at risk population taking advantage of the herd immunity. Zhang et at. used a global spatiotemporal model to project the spread of ZIKV in the Americas[6]. They estimated the posterior probabilities for time and place of ZIKV introduction in Brazil. Their work however, did not consider the effects of sexual transmission. Another work by Gao et al. incorporated both mosquito-borne and sexual transmission to investigate the spread in a deterministic manner. They concluded that sexual transmission contributes about 3.044% of the total transmission[9].

The individual-based (i.e., node-based) network model is different from basic compartmental models because it features heterogeneous mixing. An individual (i.e., a node) is only connected to a few other individuals, not the entire population. In the context of real world sexual contact networks, this is realistic. The vector population is usually very large. Therefore, it is unnecessary to use node–based models in this case. We use a homogeneous population model for the vector population. We interconnect these two populations (hosts and vectors) via the use of infection rate parameters where the infection of one population depends on the number of infected in the other population. There has been several works on epidemic spreading in interconnected networks[13]. However, due to the multi–host (one being the vector) scenario, our model is built differently from conventional interconnected networks. Our goal is to incorporate heterogeneity in the host population but reduce unnecessary computational overhead by interconnecting with the homogeneous vector population. In the host network sub-model, each host node can have either one of the four *SEIR* states (Susceptible, Exposed, Infected and Recovered). In the vector population sub-model, the population is modeled into three compartments: *SEI* (Susceptible, Exposed and Infected). We only use demography for the vector model and, the climatic variation is incorporated into the vector birth rate. This birth rate is dependent on the time/season of the year and is based on real world data collected in Miami, FL and Phoenix, AZ. The simulation model is based on Gillespie's Stochastic Simulation Algorithm (SSA)[14]. However, we have variable rate parameters in our model, which prompted us to use the non-Markovian Gillespie Algorithm (n-MGA)[15] to simulate our model which we implement using the modified version of the GEMF[16] tool.

The remainder of this paper is organized as follows. In the model formulation section we describe our model, formulate the seasonal variations, and discuss our simulation tool. The results section is divided into three subsections. In the input Data subsection, we describe the characteristics of our host contact network. In the seasonal analysis subsection, we perform several outbreak analyses using different disease starting seasons and show temporal behaviors of the disease spread and its dynamics in both host and vector populations. In the survival analysis subsection we perform extensive simulations to estimate pathogen survival due to solely sexual transmission once vector population dies out. We also compare how sexual and vectored transmission affect the epidemic lengths and epidemic sizes. The discussion section contains the summary of this paper including our key findings. Finally, in the methods section, we elaborate on our model, explain the mathematical equations of the disease model, simulation model, and the network generation process.

Model Formulation

Vector Interconnected Model

We propose a coupled network model to investigate the spread of ZIKV among human host and mosquito vector population. A small example of a coupled network is shown in Figure 1. The sexual contact network of the host population is constructed based on the sexual behavior of the human community. We use a preferential attachment model to generate that. The mosquito population is homogeneous and separate from the network; however, it is coupled to the sexual contact network and has infection paths to every host node, as any infectious vector can infect a susceptible host in its vicinity.

Each host node has four states, Susceptible (S), Exposed (E), Infected (I) and Recovered (R). The vector population is homogeneously distributed into the following three compartments, Susceptible (S_V), Exposed (E_V) and Infected (I_V). The host states are defined as probabilities and the vector compartments are defined as populations. A sum over the probabilities of being in a particular state for all the host nodes gives the total number of hosts in that state. A host node can get infected via two mechanisms, i) Bites from infected mosquitoes (at the rate λ_{HV}) or ii) Sexual contact with infected hosts (at the rate β). The elements of the adjacency matrix, a_{ij} indicates if there is a link (sexual relationship) between the nodes i and j. The parameters used in our model formulation are listed in Table 1.

Sexual Contact Network

One of the most important input to our simulation tool is the host contact network. It has been seen that sexual contact networks are not random. The degree distributions of such networks follow the scale-free structure[18]. In such networks, there are large variations in the number of sexual partners while a small group of highly active people form the core[18]. Core groups are important in sustaining pathogen transmission specially if the duration of infection, $(1/\gamma)$ is short[19].

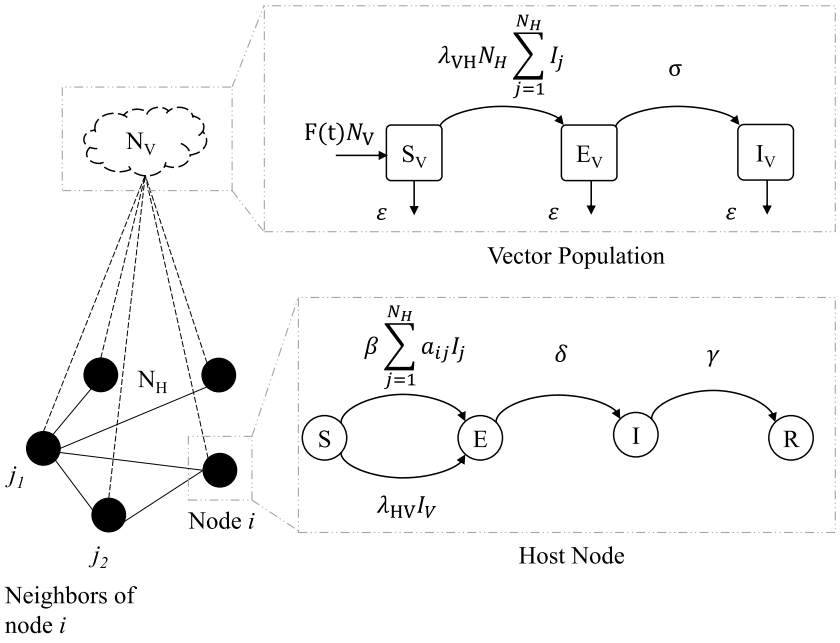

Figure 1. Vector Coupled network model for ZIKV. The solid circles indicate host nodes and the cloud shape above indicates the vector population. The solid edges connecting the nodes indicate the sexual contact network and the dashed edges indicate coupling of the vector population to all host nodes. Each host node has four states (*SEIR*) and the entire vector population is divided into three compartments ($S_V E_V I_V$).

Seasonal Variations

We intend to incorporate temporal variations in climate and the effects of season on the mosquito vectors. To simplify this task, we take the mosquito vector abundance data expressed as a fraction in the range $[0, 1]$ and multiply this with the constant mosquito birth rate. These abundance data were used in the work of Monaghan et. al.[20] which they extracted from the works of Reiskind et. al[21]. We use the data from two separate locations, Miami, FL and Phoenix, AZ. The mosquito abundance factors for these two locations over the course of 12 months are plotted in 2. The data originally had 12 sample points (monthly values). As our model requires the abundance factor on a daily basis, we use linear interpolations between data points of adjacent months.

Simulation Tool

If the transition rate parameters are kept constant, the state transitions of the host nodes are Markov processes. A stochastic algorithm such as the Gillespie SSA[14] can be used to perform simulations in most cases. The well-developed GEMFsim[16] can be used to solve such stochastic spreading processes in the networks. However, our model requires a few changes before simulations can be performed.

In our model, the transition rate parameters are not constant. The mosquito birth rate is a seasonally dependent parameter that varies according to the given vector abundance input data. The mosquito vector population is also simulated independently in parallel to the SSA transitions of the host nodes. So, the infection rates due to mosquito bites constitute a set of exogenously varying parameters. Due to these parameters, the processes become non-homogeneous Poisson. To avail this issue, we use the non-Markovian generalized Gillespie Algorithm (nMGA)[15] which assumes general renewal processes that allows exogenous variations of parameters. We modified the existing GEMFsim[16] accordingly and added a deterministic vector model ODE solver. We coupled the vector solver with the nMGA simulator to facilitate the exchanges of parameters during a simulation run. The parameters are updated after each event as the nMGA is an event-based simulation. The accuracy of this implementation depends on the time between events. The time steps should be small in order to keep the parameter discrepancy between the two coupled system at a minimum. Fortunately, it is indeed the case when infection is present in the population.

Symbol	Parameter Description	Range	Nominal Value	Reference
N_H	Total human host population	-	$1,000$	Assumed
N_V	Total mosquito vector population	-	$2,000$	Assumed
λ_{HV}	Vector to host pathogen transmission rate	-	rT_{HV}	-
λ_{VH}	Host to vector pathogen transmission rate	-	rT_{VH}	-
β	Host to host sexual transmission rate	0.00 - 0.05	0.005	Assumed
δ	Intrinsic incubation rate in hosts (day^{-1})	$1/2$ - $1/9$	$1/7$	Zhang et. al.[6]
σ	Extrinsic incubation rate in vectors (day^{-1})	$1/7$ - $1/10$	$1/8$	Zhang et. al.[6]
γ	Host recovery rate (day^{-1})	$1/3$ - $1/7$	$1/7$	Caminade et. al.[17]
$A(t)$	Mosquito abundance factor	0 - 1	See Fig 2	-
$F(t)$	Mosquito birth rate ($day^{-1}vector^{-1}$)	-	$(1/12) \times A(t)$	As defined
ε	Mosquito mortality rate (day^{-1})	$1/4$ - $1/35$	$1/12$	Gao et. al.[9]
T_{HV}	Vector to host pathogen transmission probability	0.214 - 0.8	0.634	Castro et. al.[7]
T_{VH}	Host to vector pathogen transmission probability	0.6 - 0.95	0.770	Castro et. al.[7]
r	Mosquito bite rate ($host^{-1}vector^{-1}day^{-1}$)	-	b/N_H	-
b	Mosquito bite rate ($vector^{-1}day^{-1}$)	0.4 - 0.8	0.63	Castro et. al.[7]
m	Barabási-Albert attachment parameter	1 - 10	2	Assumed
T_{end}	Simulation termination time / Max run-time (day)	-	250	Assumed
M_{start}	Simulation start month / Pathogen introduction month	1-12	11	Assumed

Table 1. Epidemic model parameters.

Results

The Input Data

The simulation tool takes as input all the parameters listed in Table 1, the initial conditions, the seasonal variation data and the contact network representation. As an initial condition, we infect (I) a single vector and assume the remaining vector population and the entire host population to be in the susceptible state (S). We specify the geographic location (either Miami or Phoenix), the starting month and the maximum run time of the simulation. For example, if we start the simulation from 1st May and run it for a maximum of 100 days for the location of Miami, the simulation will run up to 2nd week of August and it will use the abundance data of Miami, FL as input. A simulation ends before completing the maximum run-time if the epidemic dies out. As we are running stochastic simulations, we smooth out our results by averaging over 500 realizations.

An important input is the network representation of the host population. Based on human sexual behaviors, we choose the Barabási-Albert preferential attachment mechanism[22] to generate the networks. We start with a seed graph and then add nodes to it. Each new node gets connected to m existing nodes. The probability of an existing node being picked for a new connection is proportional to its current node degree. The generated network has the property that it contains a few high degree nodes that form the core group and many low degree nodes. Two such examples of scale-free networks are depicted in Figure 3.

Seasonal Analysis

We first explore the effect of seasonal variations on the epidemic progression. We use all the nominal parameter values defined in Table 1 and run simulations for both locations (Miami and Phoenix) for three distinct starting times. Taking some cues from seasonal patterns observed in the Figure 2, we choose to run independent scenarios starting on 1st of April, 1st of August and 1st of October. The maximum simulation run time, T_{end} is chosen to be 250 days. The initial condition is set as a single infected vector with everything else being susceptible. The averaged results of 500 simulations are shown in Figures 4 and 5.

We observe a large epidemic in Miami if pathogen is introduced on 1st August. In Phoenix, large epidemics occur if pathogen is introduced on both 1st August and 1st October. For all other cases, the epidemic sizes are quite small. This can be explained by comparing with the seasonal variations of Figure 2. In Miami, the mosquito abundance is high during the beginning of August. This contributes to sustaining the mosquito population longer than the other two dates in the same location (i.e. Miami). Similar arguments can be given for the two large outbreaks observed in Phoenix as well.

We also notice an interesting behavior. Despite having large positive slope in vector abundance of Miami on the month of April and large negative slope in vector abundance of Phoenix on the month of October, Miami suffers a small outbreak whereas Phoenix suffers a large one. Figure 5 shows that the infected vectors die out soon for the April epidemic in Miami although the upcoming summer sustains healthy vectors for a long time. In case of the October epidemic in Phoenix, the infected vectors rise rapidly in the first 50 days. Although after 50 days, vector population declines rapidly as well due to unsuitable climates of the upcoming winter, the initial surge in infected mosquitoes causes a large outbreak. The results indicate that the suitability of

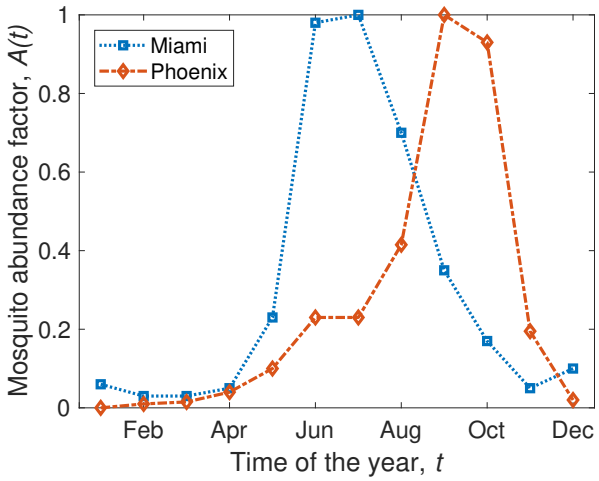

Figure 2. Variation in mosquito abundance with time. Due to changing temperature, rainfall and other climatic factors, mosquito abundance varies throughout the year. The blue dotted lines with squares show the variations in Palm Beach, FL and the orange dot-dashed lines with diamonds show the variations in Phoenix, AZ. Vector abundance peaks during June-July in Florida and during September-October in Phoenix. The Miami plot was constructed from data observed in Palm Beach, FL over 27 four-week periods from 2006-2008[2021]. The Phoenix plot was constructed from data observed in Phoenix, AZ over 10 years from 2006-2015.[2021].

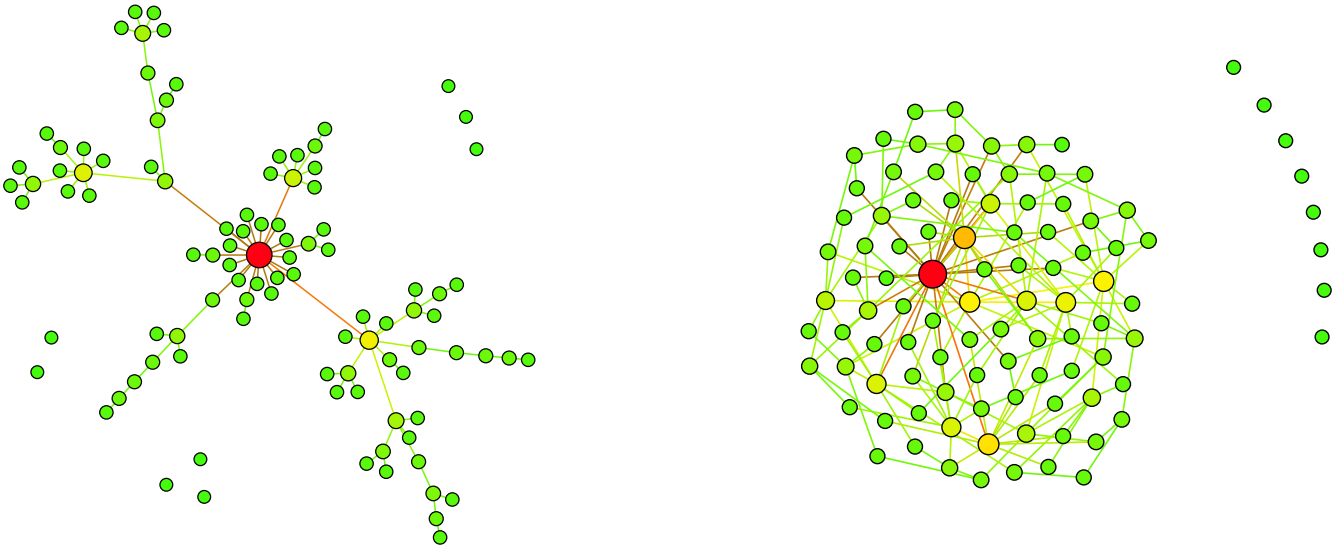

Figure 3. Barabási-Albert preferential attachment networks with 100 nodes both. We start with a Gilbert graph of 10 nodes with the edge probability $p = 0.001$. We use it as the seed and add the remaining nodes according to the preferential attachment mechanism with parameter m. The left graph is constructed for, $m = 1$ and the right graph is constructed for $m = 2$. The details of the generation process are included in the Methods section. The nodes are coded with the color spectrum from green to red which marks the nodes from low degree to high degree. Both of the graphs have scale free node degree distributions with average degree values of 1.814 (left) and 3.48 (right). The edge densities (ratio of edge count with respect to fully connected graphs) of graphs are 0.002 (left) and 0.004 (right).

the climate during pathogen introduction plays a major role in determining the size of the outbreak. The effects of changes in suitability in the following months are minimal.

Survival Analysis

Next, we explore the survival characteristics of ZIKV. To analyze survival we will be looking at epidemic attack rates, epidemic length and the pathogen survival period.

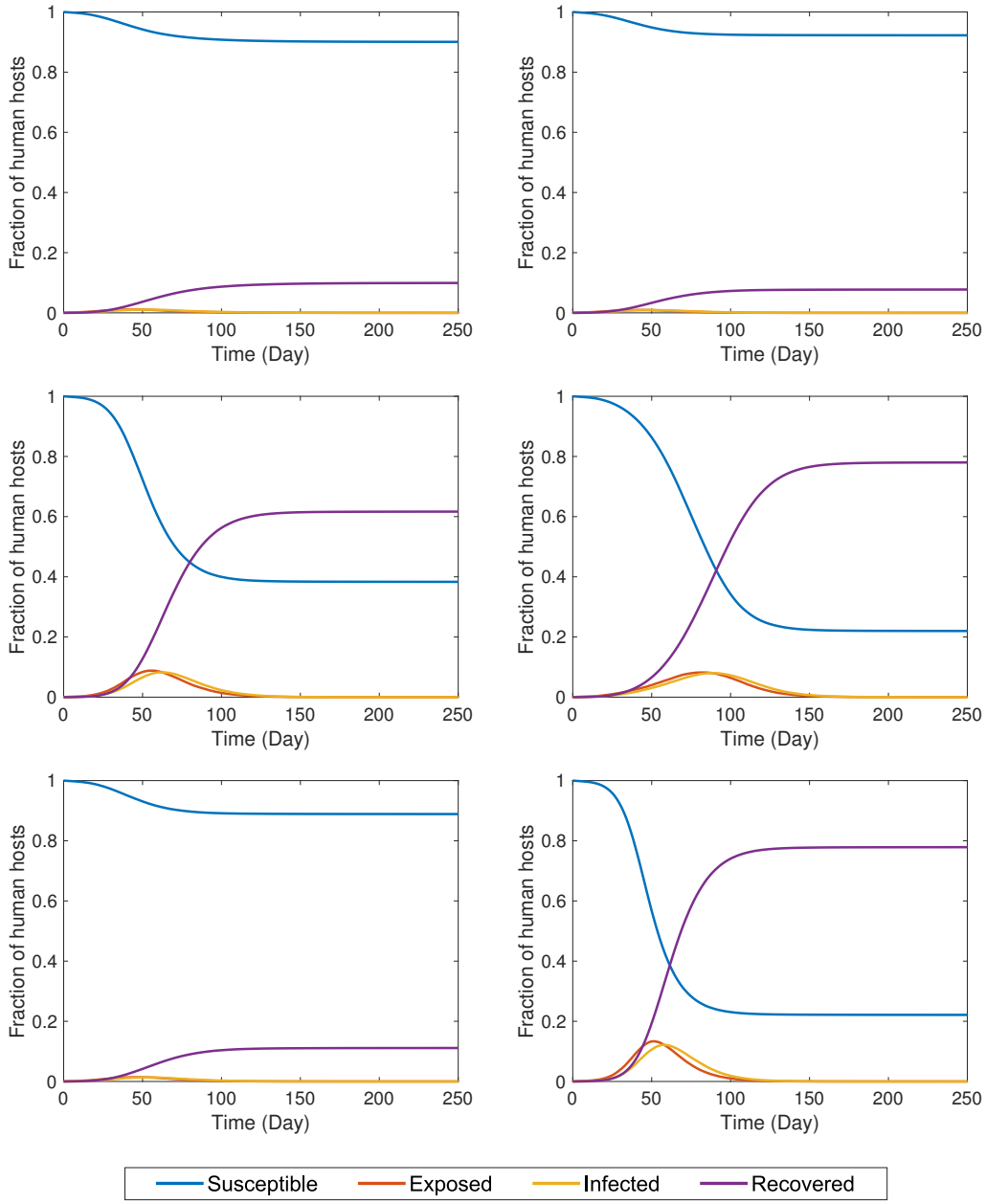

Figure 4. The time series plots of a ZIKV epidemic in the host population. Based on vector abundance data, the left column contains the results obtained for Miami and the right column contains the results obtained for Phoenix. The three rows indicate starting times of the epidemic on 1st April, 1st August and 1st October respectively. The hosts in different states are expressed as the fractions of the total population. The four states (*SEIR*) are marked by colors as indicated by the legend box in the bottom. The above plots are the averages of 500 independent stochastic simulations.

Epidemic Attack Rate (AR)

The epidemic attack rate (*AR*) is defined as,

$$AR = \frac{\textit{Number of hosts who experience infection throughout the epidemic}}{\textit{Total number of hosts, } N_H} \tag{1}$$

The value of *AR* is in the range $[0, 1]$. We sometimes also express this quantity as a percentage instead of fraction. For example, an *AR* of 0.3 means 30% of the population were infected during the epidemic and the remaining 70% never experienced any infection.

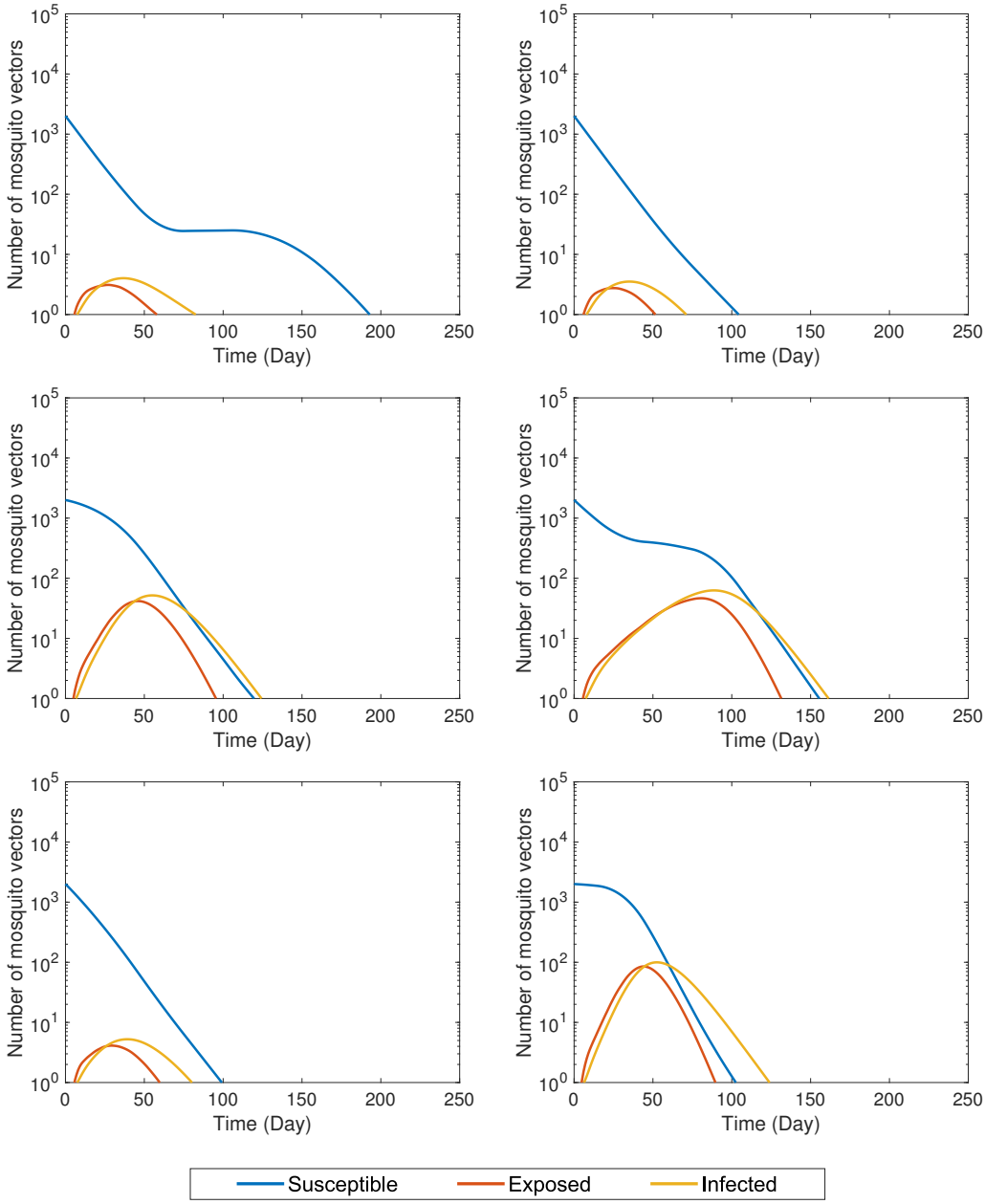

Figure 5. The time series plots of a ZIKV epidemic in the vector population. Based on vector abundance data, the left column contains the results obtained for Miami and the right column contains the results obtained for Phoenix. The three rows indicate starting times of the epidemic on 1st April, 1st August and 1st October respectively. The number of vectors in different compartments are expressed in a log scaled y-axis with respect to time. The three compartments (*SEI*) are marked by the colors as indicated by the legend box in the bottom. The above plots are the averages of 500 independent stochastic simulations.

Host Epidemic Length (T_{HL})

The host epidemic length is defined as,

$$T_{HL} = \textit{The last time instant between 0 and } T_{end} \textit{ where there is at least one infected host left} \tag{2}$$

It is the last day on the epidemic time-line an infected host can be found. In this work, when we use the term "*epidemic length*", we imply host epidemic length T_{HL} unless otherwise mentioned.

Vector Epidemic Length (T_{VL})
The vector epidemic length is defined as,

$$T_{VL} = \textit{The last time instant between 0 and } T_{end} \textit{ where there is at least one infected vector left} \tag{3}$$

It is the last day on the epidemic time-line an infected vector can be found. In typical outbreak situations, infected vectors die out before the all the infected hosts has recovered.

Pathogen Survival Period, (T_{PS})
The vector free pathogen survival period is defined as,

$$T_{PS} = T_{HL} - T_{VL} \tag{4}$$

It is a measure of how long pathogen can survive in the host population without the presence of vectors.

To see how some of the above mentioned quantities relate to our host contact network, we vary the preferential attachment parameter, m. This parameter specifies the number of connections an incoming node will create with the existing nodes during the Barabási-Albert network generation process. Effectively, this parameter controls the network density. A higher value of m will produce a well connected network with more edges, whereas a lower value will produce a sparse network. This signifies how people are engaged in sexual relationships in a community. The results obtained by varying m are shown in the top row of Figure 6. Figure 6a shows that the epidemic lengths in both Miami and Phoenix increase with the increase in network density. The pathogen survivals in both the regions also increase as evident in Figure 6b. Figure 6c shows a slow increasing trend in epidemic attack rate with the increased network density. We also see that, the attack rate in Miami is lower than Phoenix but the situation is opposite for pathogen survival. This is due to the fact that, Phoenix has a more suitable climate for mosquito vectors compared to Miami, which is reflected by the high AR found for Phoenix. Although, the pathogen survivals in both places are affected with positive correlation by m, this parameter affects Miami more than Phoenix. Relatively smaller epidemics of Miami are aided by increased sexual transmissibility resulting from higher values of m. This effect is also evident when we observe the effect of β on the same quantities.

We now focus into the effects of sexual transmission rate, β, on the epidemic length, pathogen survival, and attack rate. Increased β has a boosting effect on all three of them as evident in Figure 6 middle row. However, both the epidemic length and pathogen survival values tend to saturate after approximately $\beta = 0.035$ in Figures 6d and 6e. There are a few data points where these two quantities can even decrease with an increased value of β. This counter-intuitive behavior can be attributed to faster epidemic burnouts. The attack rate increases monotonically as expected (Figure 6f). So, even though we have a higher transmission rate, which causes a higher proportion of hosts to get infected, the epidemic peaks sooner than before and dies out faster. This is the reason why we see sudden drops in epidemic length and pathogen survival. In Figure 6e, we see a pattern similar to the one we observed before (Figure 6b) while varying m: higher sexual transmission rate extends the epidemic length in Miami more compared to Phoenix. This is due to comparatively more suitable climate of Phoenix which increases the vector epidemic length.

We now study the impact of the pathogen introduction time–i.e., the starting time of the epidemic (M_{start}). ZIKV is predominantly vector-borne which means it is heavily affected by the vector availability that changes with time. We have already shown in Figure 2 how the vector abundance can vary throughout the year. We run simulations for 12 ZIKV introduction dates (i.e., simulations start date, M_{start}) consisting of 1st day of each month of the year to evaluate how those dates affect the epidemic. The results are shown in the bottom row of Figure 6. The epidemic attack rates are directly correlated to the vector abundance as the plots (Figure 6i) closely resembles the abundance plots (Figure 2). The epidemic length plots (Figure 6g) also show somewhat positive correlation with respect to the abundance plots. We can see that epidemics have longer duration when started on the month of May in Miami and July-August in Phoenix. Notice, these peaks in the duration plots are associated with the points with maximum positive slope in the abundance plots. In accordance to our findings in the seasonal analysis, larger epidemics are also longer in general. This behavior is quite opposite to the behavior observed for higher values of the sexual transmission rate (β) where we saw that larger epidemics can die out faster. The fact is that, vectored transmission and sexual transmission affect the disease progression differently. If we analyze the pathogen survival plots (Figure 6h) we see that survival remains unchanged except for the introduction months when we have large epidemics (refer to the AR plot in Figure 6i). In those large outbreak cases (attack rate above 0.8), pathogen survival becomes negative. The negative values indicate that the simulations run out of the susceptible host pool but there are infected vectors still alive after that. The high attack rate scenarios are not realistic, as in the real world situation there will almost always be some sort of control measures in place. A proportion of the population will be alert and careful enough to avoid mosquito bites and vectored transmission. Hence, in a real world, we may never see such high attack rates.

If we consider how epidemic size (i.e., the attack rate, AR) is affected by the three parameters: m, β, and M_{start}, we see that it is least sensitive to the Barabási-Albert parameter, m (Figure 6c) compared to the other two quantities (the sexual transmission rate, β and the pathogen introduction month, M_{start}) we have varied in this analysis (Figures 6f and 6i).

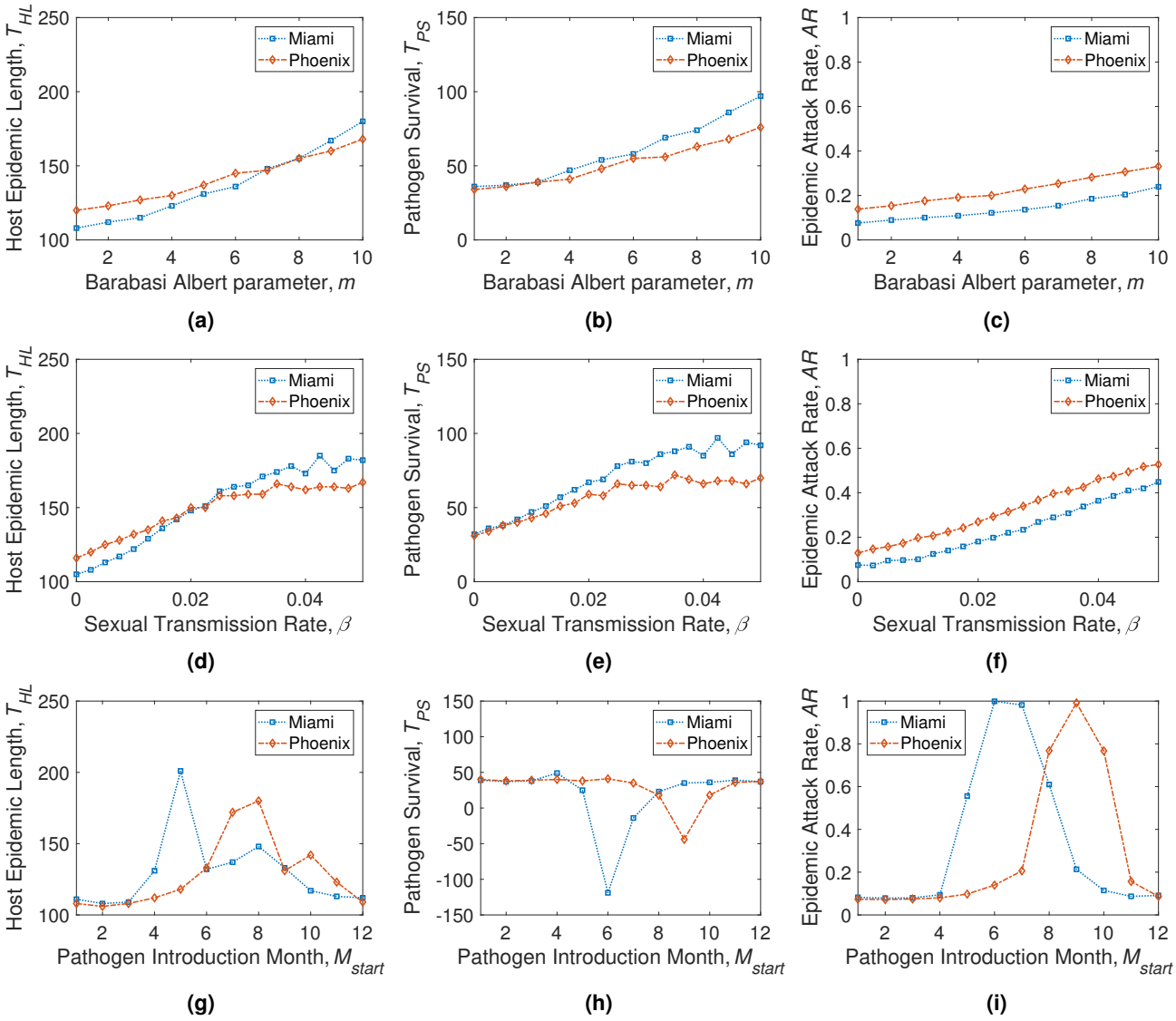

Figure 6. The plots depicting the results of pathogen survival analysis. The 1st row shows the results obtained by varying the Barabási-Albert parameter m, the 2nd row shows the results obtained by varying the sexual transmission rate β, and the 3rd row shows the results obtained by varying M_{start}, the month when the pathogen is first introduced into the population. Each data point given in the above plots is an average of 500 independent stochastic simulations.

Discussion

In this paper we have proposed a novel individual based interconnected model for ZIKV which can be used to accurately simulate vector-borne diseases which also have contact based direct transmission paths. We use heterogeneous mixing modeled by a host contact network which is interconnected to a vector population based compartmental model. We take advantage of the approximation of Non-Markovian Gillespie algorithm to simulate disease outbreaks. Once we develop the model, we use it to analyze ZIKV spreading dynamics in a population under several scenarios. First, we explore the effects of seasonal variation on the epidemic spread of ZIKV. Next, we investigate the survival potential of the ZIKA virus in the host population without vectors. In the later step we also analyze how variations of the sexual activity in a community and the transmission rate affect the disease outcome and the pathogen survival.

In our seasonal analysis, we find that epidemics are highly influenced by the environmental conditions that persist during the time of pathogen introduction. Even if climatic suitability of mosquito vectors decline rapidly after the first couple of weeks, the vectors manage to spread the pathogen in the host population rapidly and cause large outbreaks. This suggests that a ZIKV epidemic can go rapidly out of control if it is not effectively contained in the first few days.

The survival analysis provides us with a more detailed picture of the disease dynamics compared to the seasonal analysis. It is obvious from the Figures 6c, 6f, 6i and other previous works[9] that sexual transmission affects ZIKV epidemics weakly compared to the vectored transmission. However, we also see that these two transmission mechanisms affect the disease spread differently. Pathogens can survive longer due to sexual transmission when vectored transmission is relatively weak during other seasons (compared to summer). In those months, we also see that sexual transmission has a more profound effect on epidemic size. This suggests that sexual transmission could be an important aspect to consider for regions with low vector abundance. We find that with increased sexual transmission, the overall size of the epidemic will increase. However, the length of the epidemic and the pathogen survival start to saturate at some point and no longer increase monotonically. These are the scenarios where larger epidemics can also have faster burn-outs.

The conclusions drawn from our results can be useful in evaluating potential endemic scenarios for ZIKA virus disease in temperate regions. The unavailability of data on model parameters of ZIKA requires further efforts on estimation of those parameters. This model can be readily used for other vector-borne diseases with secondary transmission paths. Although Dengue virus (DENV) and Chkungunya virus (CHIKV) have not been reported to transmit sexually, the conclusions drawn in our seasonal analysis should also hold for the epidemic spreads of those diseases due to the similarities in vectored-transmission[23]. The network model used in our paper is based on sexual behavioral patterns. However, real world data on sexual contact networks, if available, can be used to generate even better predictions. Activity driven network (ADN) approach is also an effective method to model real world networks that have temporal variations. The non-Markovian Gillespie Algorithm (nMGA) which is used in this work can be readily extended to incorporate disease and vector control measures as this model supports temporally varying parameters.

Methods

Host Characterization

We use the mean-field approach developed by Sahneh et al.[24] to model spreading of pathogen among the human hosts in the sexual contact network. Our model has one single layer ($L = 1$) contact network. We define four states ($M = 4$) for a node: Susceptible (S), Exposed (E), Infected (I) and Recovered (R). We also define the probability of node i being in a state v at time t as $v_i(t)$ where v can be either S, E, I or R. For example, $I_i(t)$ is the probability that node i is infected at time t. We use the generalized epidemic mean-field simulator (GEMFsim) developed by Sahneh et. al.[16] to run our simulations. For this framework, we define the transition rate graphs in Figure 7.

The adjacency matrices corresponding to these two graphs are,

$$A_\delta = \begin{bmatrix} 0 & \lambda_{HV} I_V & 0 & 0 \\ 0 & 0 & \delta & 0 \\ 0 & 0 & 0 & \gamma \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_\beta = \begin{bmatrix} 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{5}$$

Here A_δ is the nodal transition rate matrix and A_β is the edge-based transition rate matrix. The corresponding Laplacian matrices are given by,

$$Q_\delta = \begin{bmatrix} \lambda_{HV} I_V & -\lambda_{HV} I_V & 0 & 0 \\ 0 & \delta & -\delta & 0 \\ 0 & 0 & \gamma & -\gamma \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad Q_\beta = \begin{bmatrix} \beta & -\beta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

We can derive the differential equations according to the formula developed in GEMF,

$$\frac{dv_i}{dt} = -Q_\delta^T v_i - \sum_{j=1}^{N_H} a_{ij} v_{j,3} Q_\beta^T v_i \tag{7}$$

Where a_{ij}'s are the elements of the adjacency matrix of the sexual contact network. If there is a connection between node i and node j then $a_{ij} = 1$, otherwise it is 0. Using the appropriate symbols of the compartments we get,

$$\begin{bmatrix} \dot{S}_i \\ \dot{E}_i \\ \dot{I}_i \\ \dot{R}_i \end{bmatrix} = - \begin{bmatrix} \lambda_{HV} I_V & 0 & 0 & 0 \\ -\lambda_{HV} I_V & \delta & 0 & 0 \\ 0 & -\delta & \gamma & 0 \\ 0 & 0 & -\gamma & 0 \end{bmatrix} \begin{bmatrix} S_i \\ E_i \\ I_i \\ R_i \end{bmatrix} - \left(\sum_{j=1}^{N_H} a_{ij} I_j\right) \begin{bmatrix} \beta & 0 & 0 & 0 \\ -\beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} S_i \\ E_i \\ I_i \\ R_i \end{bmatrix} \tag{8}$$

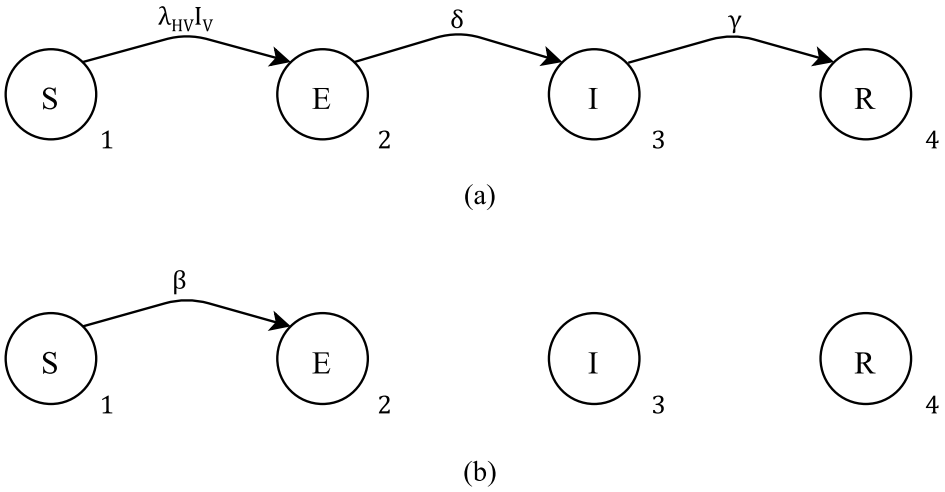

Figure 7. State transition graphs of the host *SEIR* model. The labeled nodes represent the four compartments, Susceptible, Exposed, Infected and Recovered. (a) The nodal transition rates graph. Directed link from S to E represents the infection process via mosquito bites. The rate is the product of the vector-to-host transmission rate, λ_{HV} and total number of infected vectors I_V. Directed link from E to I is the transition from the exposed to the infected state based on the intrinsic incubation period, $1/\delta$. Directed link from I to R is the curing process weighted by the recovery rate, γ. (b) Edge-based transition rate graph for the sexual contact network. Directed link from S to E represent the infection process weighted by the sexual transmission rate, β.

Our four differential equations are the following,

$$\begin{aligned}
\dot{S}_i &= -\lambda_{HV} I_V S_i - (\sum_{j=1}^{N_H} a_{ij} I_j)\beta S_i \\
\dot{E}_i &= \lambda_{HV} I_V S_i - \delta E_i + (\sum_{j=1}^{N_H} a_{ij} I_j)\beta S_i A \\
\dot{I}_i &= \delta E_i - \gamma I_i \\
\dot{R}_i &= \gamma I_i
\end{aligned} \tag{9}$$

We also have an additional equation since the sum of all probabilities would be equal to 1,

$$S_i + E_i + I_i + R_i = 1 \tag{10}$$

Therefore, we can eliminate one variable from Equation 9 and we will have to solve only the remaining three.

Vector Characterization

The mosquito vectors are modeled as homogeneous population and we use the classical Ross-Macdonald approach used by Keeling et. al. in their book[12]. We divide the vector population into three compartments, Susceptible (S_V), Exposed (E_V) and Infected (I_V). The transitions between the three compartments are showed in Figure 1. Table 1 describes the different

parameters that were used for the model. The equations for disease dynamics in mosquito vectors are given below,

$$\begin{aligned}
\dot{S_V} &= F(t)N_V - \lambda_{VH}N_H(\sum_{j=1}^{N_H} I_j)S_V - \varepsilon_1 S_V \\
\dot{E_V} &= \lambda_{VH}N_H(\sum_{j=1}^{N_H} I_j)S_V - \sigma E_V - \varepsilon_2 E_V \\
\dot{I_V} &= \sigma E_V - \varepsilon_2 I_V
\end{aligned} \tag{11}$$

We incorporate seasonality into this model by using a time dependent mosquito recruitment rate, $F(t)$. This rate depends on the time (day) of the year. The transmission parameters, the λ's are computed from the mosquito bite rate, r and transmission probability, T. The formula are given in Table 1.

Non-Markovian Gillespie Algorithm

We first describe the exact Gillespie Algorithm first for renewal processes. Then we proceed on to the nMGA (Non-Markovian Gillespie Algorithm) by Boguná et. al[15]. The following was also described on the work of Masuda et. a.[25].

We first consider N renewal processes running in parallel. Let t_i be The time elapsed since the last event of the ith process. We denote $\psi_i(\tau)$ as the probability density function of inter-event times for the ith process. The survival function of the ith process (i.e., the probability that the inter-event time is larger than t_i) is,

$$\Psi_i(t_i) = \int_{t_i}^{\infty} \psi_i(\tau) d\tau \tag{12}$$

Now, the probability that no process generates an event for time Δt is,

$$\Phi(\Delta t|\{t_j\}) = \prod_{j=1}^{N} \frac{\Psi_j(t_j+\Delta t)}{\Psi_j(t_j)} \tag{13}$$

To determine the time until the next event, Δt, we take a sample u from uniform distribution over $[0,1]$ and solve $\Phi(\Delta t|\{t_j\}) = u$. This step is computationally expensive when N is large. To improve performance, we approximate this step as proposed by Boguná et. al[15]. This approximation is exact as $N \to \infty$. When Δt is small (N is large), Equation 13 becomes[25],

$$\Phi(\Delta t|\{t_j\}) \approx exp\left[-\Delta t \left(\sum_{j=1}^{N} \lambda_j(t_j) \right) \right] \tag{14}$$

Now, the instantaneous (hazard) rate of the ith process, which is generally assumed to be a function of time since the last event is determined by,

$$\lambda_i(t_i) \equiv \frac{\psi_i(t_i)}{\Psi_i(t_i)} \tag{15}$$

With the above equations in hand, we can run the Non-Markovian Gillespie algorithm as follows,

1. Initialize t_j for all $(1 \leq j \leq N)$.
2. Determine the time until next event from,
$$\Delta t = \frac{-\ln u}{\sum_{j=1}^{N} \lambda_j(t_j)} \tag{16}$$
3. Select the process i that has generated the event with probability,
$$\Pi_i \equiv \frac{\lambda_i(t_i)}{\sum_{j=1}^{N} \lambda_j(t_j)} \tag{17}$$
4. Update the time since the last event, $t_j = t_j + \Delta t$ for all $j \neq i$. Set $t_i = 0$.
5. Repeat steps 2-4.

The original Gillespie Algorithm can be recovered from this nMGA using $\lambda_i(t_i) = \lambda_i$. This is the case for all the parameters that are constant.

References

1. Sirohi, D. *et al.* The 3.8 å resolution cryo-em structure of zika virus. *Sci.* **352**, 467–470 (2016).
2. Sikka, V. *et al.* The emergence of zika virus as a global health security threat: a review and a consensus statement of the indusem joint working group (jwg). *J. global infectious diseases* **8**, 3 (2016).
3. Mlakar, J. *et al.* Zika virus associated with microcephaly. *N Engl J Med* **2016**, 951–958 (2016).
4. Malone, R. W. *et al.* Zika virus: medical countermeasure development challenges. *PLoS neglected tropical diseases* **10**, e0004530 (2016).
5. Zika cases in the united states (2018). URL https://www.cdc.gov/zika/reporting/case-counts.html. [Accessed: 22-February-2018].
6. Zhang, Q. *et al.* Spread of zika virus in the americas. *Proc. Natl. Acad. Sci.* **114**, E4334–E4343 (2017).
7. Castro, L. A. *et al.* Assessing real-time zika risk in the united states. *BMC infectious diseases* **17**, 284 (2017).
8. Perkins, T. A., Siraj, A. S., Ruktanonchai, C. W., Kraemer, M. U. & Tatem, A. J. Model-based projections of zika virus infections in childbearing women in the americas. *Nat. microbiology* **1**, 16126 (2016).
9. Gao, D. *et al.* Prevention and control of zika as a mosquito-borne and sexually transmitted disease: a mathematical modeling analysis. *Sci. reports* **6**, 28070 (2016).
10. Testing for zika virus (2017). URL https://www.cdc.gov/zika/hc-providers/testing-for-zikavirus.html. [Accessed: 15-June-2017].
11. Kucharski, A. J. *et al.* Transmission dynamics of zika virus in island populations: a modelling analysis of the 2013–14 french polynesia outbreak. *PLoS neglected tropical diseases* **10**, e0004726 (2016).
12. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, 2011).
13. Sahneh, F. D., Scoglio, C. & Chowdhury, F. N. Effect of coupling on the epidemic threshold in interconnected complex networks: A spectral analysis. In *American Control Conference (ACC), 2013*, 2307–2312 (IEEE, 2013).
14. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal physical chemistry* **81**, 2340–2361 (1977).
15. Boguná, M., Lafuerza, L. F., Toral, R. & Serrano, M. Á. Simulating non-markovian stochastic processes. *Phys. Rev. E* **90**, 042108 (2014).
16. Sahneh, F. D., Vajdi, A., Shakeri, H., Fan, F. & Scoglio, C. Gemfsim: a stochastic simulator for the generalized epidemic modeling framework. *J. Comput. Sci.* **22**, 36–44 (2017).
17. Caminade, C. *et al.* Global risk model for vector-borne transmission of zika virus reveals the role of el niño 2015. *Proc. Natl. Acad. Sci.* **114**, 119–124 (2017).
18. Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E. & Åberg, Y. The web of human sexual contacts. *Nat.* **411**, 907–908 (2001).
19. Robinson, K., Cohen, T. & Colijn, C. The dynamics of sexual contact networks: effects on disease spread and control. *Theor. population biology* **81**, 89–96 (2012).
20. Monaghan, A. J. *et al.* On the seasonal occurrence and abundance of the zika virus vector mosquito aedes aegypti in the contiguous united states. *PLoS currents* **8** (2016).
21. Reiskind, M. & Lounibos, L. Spatial and temporal patterns of abundance of aedes aegypti l.(stegomyia aegypti) and aedes albopictus (skuse)[stegomyia albopictus (skuse)] in southern florida. *Med. veterinary entomology* **27**, 421–429 (2013).
22. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *science* **286**, 509–512 (1999).
23. Sutiono, D. R. & Gunawan, J. D. Dengue, chikungunya, and zika: Differences in similarities. *CDK-250* **44**, 176–178 (2017).
24. Sahneh, F. D., Scoglio, C. & Van Mieghem, P. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Netw.* **21**, 1609–1620 (2013).
25. Masuda, N. & Rocha, L. E. A gillespie algorithm for non-markovian stochastic processes. *arXiv preprint arXiv:1601.01490* (2016).

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Science Foundation under Grant Award CIF-1423411, United States Department of Agriculture under the Grant No. 2015-67013-23818 and 3020- 32000-008-04-S.

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Additional information

To include, in this order: **Accession codes** (where applicable); **Competing financial interests** (mandatory statement).

The corresponding author is responsible for submitting a competing financial interests statement on behalf of all authors of the paper. This statement must be included in the submitted article file.