A spatio-temporal individual-based network framework for West Nile virus in the USA: parameter estimation and spreading pattern selection using approximate Bayesian computation

by

Sifat Afroj Moon

B.S., Bangladesh University of Engineering and Technology, 2013

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Electrical and Computer Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Caterina M. Scoglio

Copyright

Abstract

West Nile virus (WNV) —a mosquito-borne arbovirus— entered the USA through New York City in 1999 and spread to the contiguous USA within three years while transitioning from epidemic outbreaks to endemic transmission. The virus is transmitted by vector competent mosquitoes and maintained in the avian populations. WNV spatial distribution is mainly determined by the movement of residential and migratory avian populations. We developed an individual-level heterogeneous network framework across the USA with the goal of understanding the long-range spatial distribution of WNV. To this end, we proposed three distance dispersal kernels model: 1) exponential —short-range dispersal, 2) power-law —long-range dispersal in all directions, and 3) power-law biased by flyway direction —long-range dispersal only along established migratory routes. To select the appropriate dispersal kernel we used the human case data and adopted a model selection framework based on approximate Bayesian computation with sequential Monte Carlo sampling (ABC-SMC). From estimated parameters, we find that the power-law biased by flyway direction kernel is the best kernel to fit WNV human case data, supporting the hypothesis of long-range WNV transmission is mainly along the migratory bird flyways. Through extensive simulation from 2014 to 2016, we proposed and tested hypothetical mitigation strategies and found that mosquito population reduction in the infected states and neighboring states is potentially cost-effective.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, it's my pleasure to express my most sincere gratitude and appreciation to my thesis adviser Dr. Caterina Scoglio for her continuous guidance, encouragement and kind support. Then I thankfully acknowledge the immensely helpful guidance from Dr. Lee Cohnstaedt. Their thoughtful comments and suggestions at every stage of my research were simply appreciable and essential. I also would like to express my gratitude to Dr. Don Gruenbacher for being my committee member and for his unconditional supports as well. I can't but grab the opportunity to express my sincere gratitude to my fellow lab-mates in the NetSE Group for their thoughtful discussions, assistance, and for all the quality time we spent in the entire time.

Most of all, I am grateful to my parents for their encouragement, sacrifice and inspiration. Finally, I would like to thank my husband, who is my friend, colleague, lab-mate, and life partner. This accomplishment would not have been possible without his support.

Chapter 1

Introduction

West Nile disease (WND) is a vector-borne zoonosis which may result from infection by West Nile virus (WNV), a member of the family Flaviviridae, genus Flavivirus. This virus is the most common cause of arboviral disease in the United States.[4] From 1999 to 2017, more than 48 thousands WNV disease cases were reported to the Centers for Disease Control and Prevention (CDC) and more than two thousands of these reported cases resulted in death.[1] WNV is maintained in an enzootic transmission cycle between competent mosquitoes and birds. Birds are the reservoir and amplifying host for this virus. The US Centers for Diseases Control and Prevention (CDC) has identified WNV infection in more than three hundred species of birds. Infected bird movement is likely a key factor that affects the geographic spread of WNV, especially given the different habitats and routes of various species. Although many bird species may be infected with WNV, the American robin is considered an important amplifier of WNV and maybe a driver geographic spread because WNV-infected American robins have low mortality and high viremia.[5;6] Members of the *Culex* genus of mosquito are the principal vectors of this virus in the United States.[7] Humans, horses, and other mammals can be infected with WNV. However, these infections result in relatively low virus titers (viremia) therefore the infected animals and people are considered dead-end hosts (not capable of infecting feeding mosquitoes). Therefore, they do not have any epidemiological impact on WNV transmission or geographic spread.[8]

To understand the transmission dynamics of WNV, several mathematical models have been developed.[5;9–12] These models predict the threshold conditions for WNV spreading in different scenarios. However, most of these models do not consider the spatial dynamics of WNV. Space or geographic spread has a significant role in WNV disease dynamics and modeling of WNV spatial spreading is complex because of the interactions of multiple potential mosquito vectors, avian amplifiers, and mammalian hosts. Liu et al.[11] developed a patchy model to analyze the spatial spreading of WNV, where patches are geographical space. They assumed patches are identical, spatial dispersal of birds and mosquitoes are symmetric within patches, and movement of birds and mosquitoes are only one-dimensional. According to this investigation, long-range dispersal of infected bird populations determines the spatial spread of WNV, not the dispersal of infected mosquito populations. Other investigators proposed a reaction-diffusion model[12], where they have spatially extended the non-spatial model of Wonham et al.[10] to mathematically estimate the spread of WNV. Here, diffusion terms in the reaction-diffusion partial differential equations represent vector mosquito and host bird population movements. They identified traveling wave solutions in their model and calculated the rate of spatial spread of infection. Durand et al.[13] developed a discrete time deterministic meta-population model in order to analyze the circulation of WNV between Southern Europe and West Africa. Another spatial model proposed by Maidana and Yang[14] used a system of partial differential reaction-diffusion equations. They also calculated the speed of disease dissemination by investigating the traveling wave solution of their model. They concluded, mosquito movements do not play an important role in disease dissemination. In addition, they included vertical transmission in their model and determined that vertical transmission is not an important factor for the spatial spread of WNV.

Most WNV spread models are mathematical deterministic compartmental models. However WNV spread is highly stochastic because of the demography and movement of hosts and vectors varies between different locations. The major weaknesses of these models are the number and complexity of the compartments required to account for the many host and vector populations. In turn, the number of compartments increases the number of unknown parameters. Approximation of these parameters in any biological system is very challenging

and prone to estimation errors which can create inaccuracies in the model outputs.

We developed an individual-based heterogeneous network framework to understand WNV geographic spread. To build the network framework, we used the American Robin population density across the contiguous United States. The demographic characteristics of avian host populations and vector populations are not homogenous geographically, so we used a heterogeneous network framework. The transmission intensity of WNV depends on the abundance of WNV-infected vector mosquitoes in a given location. Mosquito population numbers fluctuate with local weather and season throughout the year, therefore we used a temperature dependent transmission rate. Although dead-end hosts cannot spread WNV to mosquitoes, we have quantified WNV case data only for humans, which we used to estimate unknown parameters.

To understand the WNV spatial distribution, we proposed distance dispersal kernels, which describes the probability of dispersal with respect to distances. In this framework, we proposed three types of distance dispersal kernels: 1) exponential, 2) power-law, and 3) power-law biased by flyway. Then we compared the three distance kernels using approximate Bayesian computation based on sequential Monte Carlo sampling (ABC-SMC) method.[15–20] After conducting an extensive simulation for 2014-2016, we observed that an adapted fat-tailed or power-law kernel, which has long-distance links in specified directions can best describe the WNV human case data. We tested this network framework for the best kernel with the human case data and found that simulated results for more than 41 states of 49 states are consistent with the reported WNV cases. We proposed several theoretical mitigation strategies to control WNV and calculated their estimated costs. From the analysis of mitigation strategies, we suggest that potentially effective mitigation policies would include the application of mitigation control in areas with active transmission and in immediate neighboring states.

Chapter 2

Materials and methods

In this section, we present data sources, an epidemic model for WNV, then develop a network framework for WNV geographic spread across the United States. At the end of this section, we present a statistical tool, approximate Bayesian computation with sequential Monte Carlo sampling (ABC-SMC) for parameter estimation and model selection.

2.1 Data

The study area of this research was the contiguous United States where WNV is considered endemic. We modeled WNV case distributions for 2014-2016. We used three data sets each year to develop our model. The first dataset contained the average monthly temperatures. Mosquito vector abundance correlated with temperature. Temperature data was from the National Centers for Environmental Information.[2] The second dataset contains American Robin population data from *eBird*.[3] This is a database for bird abundance and distribution, which is formed by the Cornell Lab of Ornithology and National Audubon Society. We used total observation of American Robin in each state of the USA for each month. The robin data set was used to train the network model. The American Robin is abundant throughout the United States and is a preferred food source for many WNV-competent mosquito species.[21] Based on host feeding patterns of the *Culex* genus of mosquitoes, robins are the most common

WNV amplifying host.[22–24] Other important susceptible birds, such as American crow were not used because although they are an indicator species (high crow mortality), they are unlikely to spread virus geographically as they are mostly a residential species. In addition, as an indication of epidemic start point, we used WNV human incidence data. Many species of birds have long-distance migration during the spring and fall. Therefore the network does not focus on one long-distance migrating bird species but aggregates all species along the known flyways. To estimate model parameters we used human case data for WNV from CDC,[1] which is the third dataset.

2.2 WNV Epidemic model

To explore WNV long-distance spatial distribution in the USA, we used an individual-based heterogeneous network framework. In this framework, birds are on the individual level, a node represents an individual bird and connection between nodes is the possibility of virus dispersal from one infected bird to another susceptible bird by mosquito vectors. Links or connections are formed by movement of birds or movement of vectors. If there is no link between nodes then infected birds and insects are not moving virus between nodes. All virus transmission occurs by local competent vector mosquitoes. There is some evidence of bird-to-bird transmission, but it likely does not contribute to or maintain outbreaks. We split the bird population into four compartments; susceptible, exposed, infected, and recovered. Although, in the literature most mathematical models do not consider the exposed avian class when modeling WNV.[10;14;25;26] Birds transmit virus to mosquitoes when a susceptible mosquito vector takes an infected blood meal, then the mosquito becomes infectious after the extrinsic incubation period (EIP), or the time needed for the virus to spreads from the mosquito mid gut to the salivary glands; usually this process takes 7 to 14 days.[5;27] In addition, an infected bird can infect many mosquitoes simultaneously and also an infected mosquito can bite many susceptible or infected birds. Therefore, there is some delay in the system, to represent this delay we added the exposed class. We estimated exposed period from data by using the approximate Bayesian computation with sequential Monte

Carlo sampling (ABC-SMC) method. After the exposed period, birds entered the infected compartment and an infected bird transitions to recovered after 4-5 days. To simulate this model, we used generalized epidemic mean-field (GEMF) framework developed by the Network Science and Engineering (NetSE) group at Kansas State University.[28] In GEMF, each node stays in a different state and the joint state of all nodes follows a Markov process.[28–30] The node level description of this Markov process is:

$$Pr[x_i(t+\Delta t)=1|x_i(t)=0,X(t)]=\beta(T)Y_i\Delta t \tag{2.1}$$

$$Pr[x_i(t+\Delta t)=2|x_i(t)=1,X(t)]=\lambda\Delta t+o(\Delta t) \tag{2.2}$$

$$Pr[x_i(t+\Delta t)=3|x_i(t)=2,X(t)]=\delta\Delta t+o(\Delta t) \tag{2.3}$$

Here, $X(t)$ is the joint state of all individual nodes at time t. $x_i(t)$ is a node state, $x_i(t) = C$ means node i is in C compartment at time t, $C = 0, 1, 2, 3$ corresponds to susceptible, exposed, infected, and recovered compartment. Y_i is the number of infected neighbors of node i, $\beta(T)$ is the probability of transmission from one infected bird to one susceptible bird, which is a function of temperature, λ is the rate for exposed to infectious state, and finally, a node recovers from infectious state at a rate δ.

2.2.1 Zoonotic spillover transmission

To model disease transmission from the bird population to human population, we added a zoonotic spillover transmission compartment. We modeled occurrence of human cases as a Poisson process.[26;31] This part of the framework can be expressed as the following equation:

$$\Delta Ih_{n_s}=Poisson(\eta Y_{n_s}) \tag{2.4}$$

In this equation, Ih_{n_s} is number of infected human cases at n sub-network in s time steps, where $s = 1, 2, 3.....$ are the discrete time steps, Y_{n_s} is infected bird population in sub-network n, and η is a scaler quantity, accounts for the contact rate and probability of pathogen

transmission from bird to human. We calculated WNV spilling over to humans by using a Poisson random number generator.

2.2.2 Temporal transmission rate and environmental conditions

The transmission rate for WNV is sensitive to weather data as mosquito abundance depends on the environmental conditions. Temperature, precipitation, landscape features, daylight conditions etc. are environmental conditions, which has an impact on the transmission dynamics of WNV.[32] In this research, we considered average monthly temperature data, optimal mosquito season,[33] and suitable temperature range for co-occurrence of WNV and competent mosquito species. Temperature plays a very important role in the transmission dynamics of WNV because mosquito longevity and EIP are sensitive to temperature. Mosquito longevity and EIP decrease with the increase of temperature. However, there is no straightforward relationship of vectorial capacity for WNV with temperature. If incubation period decreases more than longevity, then mosquitos will be infective longer. However if longevity decreases more than incubation period, then mosquitos will not be able to transmit the virus. We used information about rainfall in this research implicitly through optical mosquito season. Optimal mosquito season of any location is estimated from monthly average temperature and rainfall data for that location.[33] In this model, we used a simple linear relation of transmission rate with temperature in a temperature window from 12°C to 32°C in the optimal mosquito season. Outside this window, transmission rate is very low. Suitable temperature for co-occurrence of WNV and *Culex pipiens* is around 12° to 27°C and for *Culex quinquefasciatus* is 20°C to 32°C.[33] Survival rate to adult stage for *Culex quinquefasciatus* is significantly high when temperature is in 20°C to 30°C.[34] For *Culex tarsalis* favorable temperature for WNV development start after 14°C,[35] however larval survival reduced after 30°C temperature.[36] To compute the transmission rate of any link from node a to node b, we used temperature of the location of node b. Transmission rate for a location l is, $\beta_l(T) = \beta_\circ(T_{lm} - T_\circ)$; here, $\beta_\circ$ is the proportional constant, what we estimated by using ABC-SMC method, T_{lm} is the average temperature for month m in location l and $T_\circ$ is the

threshold temperature. Threshold temperature for this model is 12°C. As the temperature is space dependent, our transmission rate also differs across the network. This individual-level heterogeneous network model gives us this flexibility to use different transmission rate at a time for different parts of the network.

2.3 Network framework

For the spatial dynamic characteristics of WNV transmission, we built a network framework, which has 49 sub-networks one for each adjoining states of the contiguous United States plus the District of Columbia. The number of nodes in each sub-network is proportional to the size of the avian population in that state.[3] We considered the mosquito season June-October for the simulation period. Although the mosquito season is not the same for all states, mosquitoes are active from June to September in all of the states at these times.[33]

The network for the avian population is (V, E). Here, V is the set of nodes, which is the union of nodes of all sub-network, $V = SN1 \cup SN2 \cup SN3 \cup ..... \cup SN49$, here SNi is a set of nodes in the sub-network i and E is the set of links among individual nodes. To build sub-networks, we used the total number of observations of American Robin for states per month in the simulation time period. $|SNi| = \max_{mj=m1:m2}(OBS^i_{mj}) * S_c + N_0$, here, OBS^i_{mj} is the total number of observations of American Robins in state i in month mj, N_0 is the error term and $N_0 \sim N(5, 2)$ for this model. $m1$ is the first month after May and $m2$ is the last month before October when the average monthly temperature is greater than T_0. S_c is the scaling constant.

In each sub-network, we assumed that nodes are connected through *Erdos-Renyi (n,p)* random network topology.[37] In this network topology, we created links randomly among nodes with a probability p. Here, n is the number of nodes in a sub-network and p is the probability to form an edge. We set the probability $p = R * log(n)/n$, here R is a constant ($R \geq 2$), as this value is more than the threshold value for the connectedness of an *Erdos-Renyi* graph,[38] so nodes of a sub-network are locally connected. We will refer these networks as a local network in the subsequent sections of this paper. To build connections among sub-networks, we

considered long-distance dispersal kernels, which describe the probability of dispersal with respect to distances. Dispersal kernels provide a simple model of dispersal to model dispersal events. For long-distance events, we used three types of kernel models; 1) Exponential, 2) power-law, and 3) power-law-flyway, which is a power-law kernel biased by flyway. A simple caricature of the network is shown in Fig. 2.1. There are three sub-networks, A, B, and C. The links, which formed local networks are shown by solid lines. These links are introduced by *Erdos-Renyi (n,p)* network topology. Dashed lines are inter-links among sub-networks. These links established by using long-distance dispersal kernels.

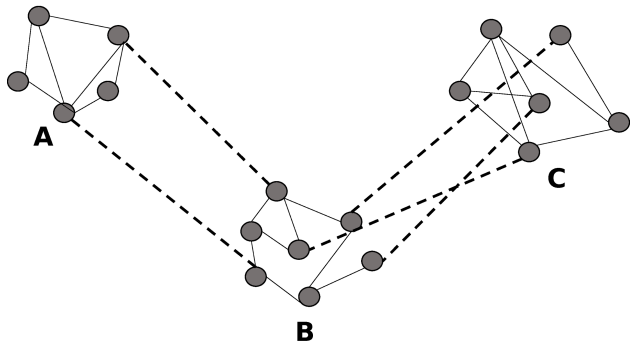

Figure 2.1: **A simple caricature of the actual contact network for the avian population.** Here, A, B, C are three sub-networks. Solid lines represent intra-links in a sub-network and dashed lines represent inter-sub-network links.

2.3.1 Exponential distance kernel

In this distance kernel, connection probability among sub-networks will decrease exponentially with distance. Probability to form a link is:

$$P(d_{ij}) = K_e * \exp(-K_e * d_{ij}) \tag{2.5}$$

Here, d_{ij} is the distance between sub-network i and j, K_e is the shape parameter of exponential distribution kernel. For distance between two states, we took the distance between their centroids. The network with the exponential dispersal kernel was created as follows:

Step 1 Calculate the distance among sub-networks. d_{ij} is the distance between sub-

network i and j.

Step 2 Calculate $P(d_{ij})$, this is the probability to form a link between sub-network i and j.

Step 3 Generate a random number $rand$ for each pair of nodes (a,b), where $a \in i$ and $b \in j$.

Step 4 If $rand < P(d_{ij})$ then an undirected link will form between node a and b.

Inter-links among sub-networks, generated by exponential distance kernel are shown in Fig. 2.2a.

2.3.2 Power-law distance kernel

Power-Law, heavy-tailed, or fat-tailed distribution allows occasional long-range transmissions of infection with frequent short-range transmissions. In this fat-tailed distance kernel, there is a greater chance of creating links over the same long-distances compared to the exponential kernel. Power-law transmission kernel was used previously to model spatial dynamics of several infectious diseases, for example, in plant epidemiology,[39] in 2001 foot-and-mouth disease epidemic,[40] and also, in human diseases.[41] In power-law connections,[42] the probability of connectivity among sub-networks will decrease with distance according to the following equation:

$$P(d_{ij}) = (K_{pl} - 1)/d_{min} * (d_{ij}/d_{min})^{-K_{pl}} \tag{2.6}$$

Here d_{min} is minimum distance among sub-networks and K_{pl} is the power-law parameter. The process to build this network is similar to a network for exponential kernel with the only difference being the calculation of $P(d_{ij})$. Inter-links among sub-networks for power-law distance kernel are shown in Fig. 2.2b.

2.3.3 Power-law distance kernel biased by flyway

To form this distance kernel, we included the migratory behavior of birds. Migratory birds can spread pathogens during the migration periods.[43;44] According to the United States Fish and Wildlife Services and Flyway Councils, there are four flyways in the United States; the Atlantic flyway (AF), the Mississippi flyway (MF), the Central flyway (CF), and the Pacific flyway (PF).[45] Although flyways overlap and the migratory patterns are very complex, these migratory routes play a vital role in the long-distance spreading of WNV.[46] To build this distance kernel, we considered two types of links among sub-networks; 1) links which are formed for residential or short-distance migratory bird movements and 2) links which are formed for long-distance migratory bird movements. For the first type of links, we used an estimated movement range of 500 km,[47] these connections are unrelated to flyways. For the second type of connections, we considered two migration periods; spring migration (April - June) and late summer/fall migration (July - September);[30] during the spring migration, we established long links from south to north and in late summer/fall migration, the reverse. To establish any long link, we picked two sub-network and establish a link if they were in the same flyway with probability $P(d_{ij})$ (Eq. 2.6), these links were directional and direction was imposed with respect to migratory period. Inter-links among sub-networks for this kernel were shown in Fig. 2.2c. The algorithm to create this network was:

Step 1 Calculate the distance among sub-networks. d_{ij} is the distance between sub-network i and j.

Step 2 Calculate $P(d_{ij})$ using Eq. 2.6, this is the probability to form a link between states i and j.

Step 3 Generate a random number $rand$ for each pair of nodes (a,b), where $a \in i$ and $b \in j$.

Step 4 If $rand < P(d_{ij})$ and $d_{ij} < 500km$ then an undirected link will form between node a and b.

Step 5 If $rand < P(d_{ij})$ and $d_{ij} > 500km$ and states i and j are in the same flyway then an directed link will form between node a and b according to the migration period.

2.3.4 Temporal network behavior

Bird populations are not constant in any region, they change with time because of bird movement. To consider this fact, this study adds a node property, namely, *Activity*. This property can hold two values: $1 = Active$ and $0 = Inactive$. In the entire network, only *Active* node can contribute to the spreading of the WNV. By controlling this property, we varied the size of the active node population in any sub-network with respect to the variation of the avian population in that region. The length of the simulation each year was five months (June - October). Then, each month nodes are activated randomly according to the total number of birds observed in that region in that month.

2.4 ABC-SMC for parameter estimation and model comparison

In this framework, we adopted approximate Bayesian computation based on a sequential Monte Carlo sampling (ABC-SMC) method for parameter estimation and model selection.[15–20]

2.4.1 Parameter estimation

ABC-SMC is a computational method of Bayesian statistics that combines a particle filtering method with summary statistics. This method is ideal for a stochastic complex model where likelihood function is intractable or computationally expensive to evaluate. ABC estimates the posterior distribution of parameters from data. Let, θ is a parameter vector to be estimated. The goal of the ABC is to approximate the posterior distribution, $\Pi(\theta|d) \propto f(d|\theta)\Pi(\theta)$, where prior distribution of parameters $\Pi(\theta)$ are given and $f(d|\theta)$ is

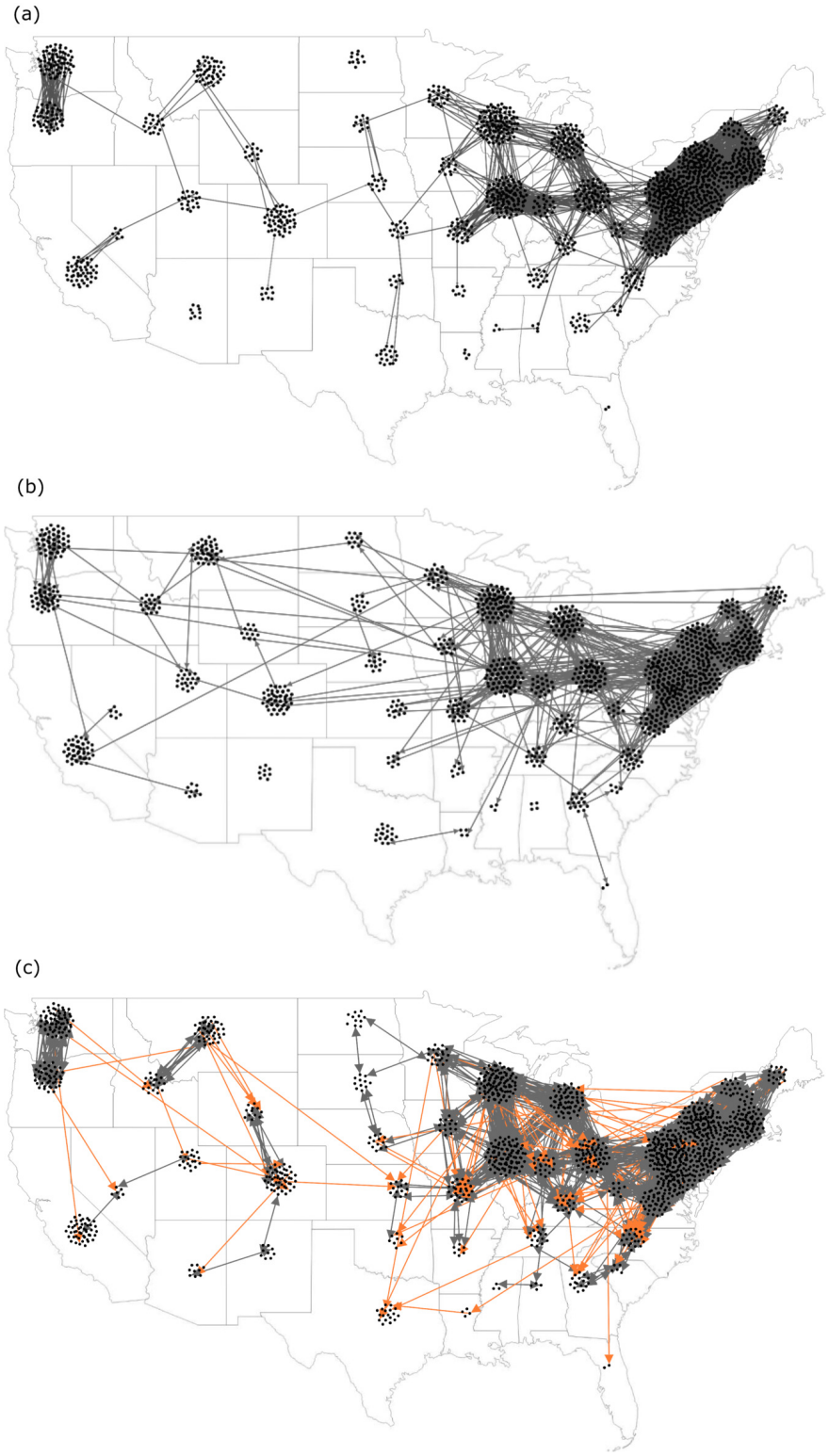

Figure 2.2: **Inter-links among sub-networks;** a) for exponential distance kernel, b) for power-law distance kernel, and c) for power-law distance kernel biased by flyway. Gray links represent undirected links and orange links represent directed links (for spring migration –northbound; for late summer/fall migration –southbound). Intra-links are not visible here. These are one realization of the stochastic networks, which are rescaled by 0.1 for better visualization.

the likelihood of θ given the data d. This method samples parameter values from their prior distribution through subsequent SMC rounds. Intermediate distribution of the parameter is $\Pi(\theta|dist(x,d) \leq \epsilon_i); i = 1, 2, ...P$. The target posterior distribution is $\Pi(\theta|dist(x,d) \leq \epsilon_P)$. Here, x is the simulated data set, $dist$ is the distance function, ϵ is the tolerance and P is the number of SMC rounds or the number of populations, where $\epsilon_P < ..... < \epsilon_2 < \epsilon_1$.[48] This is an adapted sequential importance sampling. In each SMC round, it uses perturbation kernel to sample a parameter set. After each simulation of the model, the model output and data are compared using some goodness-of-fit metrics. A parameter set is accepted if the distance between the model output and data is less than the tolerance level. The accepted parameter set is a particle and accepted particles form a population for that SMC round. We used two goodness-of-fit metric or distance function in this research. The first goodness-of-fit metric is squared root of the sum of squared error between observed incidence data and simulated incidence data for any proposed parameter set. The first goodness-of-fit metric for this model is:

$$dist_1(x,d) = \sqrt{\sum_{i=1}^{w}\sum_{j=1}^{s}(x(i,j) - d(i,j))^2} \tag{2.7}$$

Here, *x(i,j)* is simulated incidence model data for i week and for j location. The second goodness-of-fit metric is the absolute difference between the number of infected states from observed data and simulated data, infected state defined as a state where at least one infected individual has reported. The ABC-SMC algorithm, we adopted for this model from Toni et al.,[15] which has given in Appendix B. We used this algorithm separately for estimating parameters for this three distance dispersal kernel network models. As our models are an event based stochastic simulation, we simulated them 30 times with GEMF for each particle to get 30 realizations of the system. Then we take the average of these realizations. As the average over the multiple runs of a stochastic system holds more information than a single stochastic run.

2.4.2 Model comparison

In many areas, researchers deal with model selection. Bayesian theory is a comprehensive method to make inference about models from data. Approximate Bayesian computation was used in many research areas for model selection.[49] To compare among three distance kernels, this investigation used ABC-SMC model selection framework.[15;50;51] For given data d, the marginal posterior probability of model m is:

$$Pr(m|d) = Pr(d|m)Pr(m)/Pr(d) \tag{2.8}$$

Here, $Pr(d|m)$ is the marginal likelihood and $Pr(m)$ is the prior probability of the model. We used a uniform distribution for prior distribution of unknown parameters. For each model, we have four unknown parameters; network parameter K (K_e is the network parameter for the exponential kernel and K_{pl} is the network parameter for the both power-law kernels), constant for transmission rate β_0, transition rate from exposed to infectious state λ, and zoonotic transmission spillover rate η. In each population, we took 1000 particles. We used Bayes factor to compare a model with another model. For model m_i and m_j, Bayes factor[52] is,

$$B_{ij} = \frac{Pr(m_i|d)/Pr(m_j|d)}{Pr(m_i)/Pr(m_j)}, \tag{2.9}$$

Here, $Pr(m_i)$ is the prior and $Pr(m_i|d)$ is the marginal posterior distribution of model m_i. The Bayes factor is a summary evidence in favor of one model over another supported by the data. If B_{ij} is in range 1-3, we can conclude that summary of the evidence against mj in favor of m_i is very weak. If B_{ij} is in range 3-20, we can conclude that summary of the evidence against m_j in favor of m_i is positive.[52] The ABC-SMC model selection algorithm is very similar to the algorithm for parameter estimation. Here, m is the model indicator, $m \in 1, 2, ..., M$, M is the number of model. In this research, we had three network models ($M = 3$) to compare.

m = 1: exponential kernel network model,

m = 2: power-law kernel network model, and

m = 3: power-law kernel influenced by flyway network model.

In each population, the model selection algorithm starts by sampling the model parameter m from the prior distribution $\Pi(m)$. Then the algorithm proposes a new set of parameters (particle) from the sets of parameters of the model m from the previous population. The Bayes factor was calculated from the final population of m. The algorithm for model selection has given in Appendix B. Although ABC-SMC is an accurate statistical tool for parameter estimation and model selection, however, the results of this method are sensitive to summary statistics.[53] For our case, no summary statistics were required because we used the entire set of data and we compared the simulated and observed dataset directly by using goodness-of-fit or distance metric. A full dataset is sufficient to get the consistent result from approximate Bayesian Computation.[54]

2.5 Mitigation strategies

The role of mosquito populations in WNV transmission is expressed by disease transmission rate β. This framework used different transmission rates in different parts of the network corresponding to the local mosquito abundance. Using this heterogeneous feature in the framework, we evaluated theoretical mosquito population management measures to reduce the outbreak size or transmission rates in the state level. Some states such as Kansas, do not have statewide mosquito surveillance or management, but in these theoretical scenarios, it is assumed they can develop or benefit from effective statewide mosquito management programs. The framework will simply estimate how much the mosquito abundance is reduced or maintained based on the theoretical outcomes of coordinated control. Furthermore, we realize mosquito control is generally conducted on a county or municipal level, but the human case data is only available on a state level. Therefore the recommendations are for the lowest resolution of the data, which is state level but applies to counties and municipalities as well. If vector management is increased in a sub-network, then transmission rates will be changed by, $\beta_r = \frac{\beta}{RF}$, here β_r is the reduced transmission rate and RF is the reduction factor. Then management costs will be $Cost = RF * NS_c$, here NS_c is the number of states

where control measures were applied. We considered supplemental management measures with the existing management measures. We used two types of mitigation strategies across the United States, 1) dynamic infected place tracing strategy and 2) static ranked based strategy.

In the infected place tracing, we traced the infected states, then plan the mitigation strategies according to them. For this type of mitigation strategies, we considered three cases; 1) *case-1: only infected*: applied control only in the infected states; 2) *case-2: infected & first neighbors*: applied control in the infected states with its first neighboring states (whose distance is less than $500km$), and 3) *case-3: infected & first neighbors & second neighbors*: applied control in the infected states with its first neighboring states, and also with its second neighboring states (whose distance is in $500 - 1000km$). For infected tracing control measure, we kept track of infected places monthly. If SNi sub-network is infected for month t, then control measures were applied for the month $t + 1$ based on these three cases.

In the static ranked based mitigation strategy, we ranked the states by different variables (for example, temperature, size of the avian population etc.). For this strategy, we considered three cases; 1) *temp.*: states ranked by temperature, 2) *pop.*: states ranked by avian population size, and 3) *temp. & pop.*: states ranked by temperature and avian population size both, then we applied management measures in the top 30% of the states.

Chapter 3

Results

We developed a novel flexible individual based heterogeneous network framework to test three WNV dispersal kernels across the contiguous United States based on human case data distributions. We used this framework for the year 2014, 2015, and 1016. The results for network formulation, parameter estimation, and dispersal kernels selection using Bayesian inference are given below for the year 2015 and the results for other two years are given in the Appendix A.

3.1 Network framework

In this spatial-temporal individual-based heterogeneous network framework, we used three distance kernel models. The fundamental basic WNV epidemic model is the same for all the three network kernels. In the entire network, there are 49 sub-networks representing the 48 adjoining contiguous states plus the District of Columbia. All sub-network nodes are locally connected. The topology of the local network is *Erdos-Renyi*. The total nodes for the year 2015 was $|V| = 7657$ and the scaling constant is $S_c = 0.02$. Here, $E = E_l \cup E_{dd}$; $|E_l|$ is the number of total intra-links for all local networks, which is around 167000-170000 and $|E_{dd}|$ is the number of total inter-links among sub-networks. The description of sub-networks is provided in S4 Table in the Appendix C. We started the epidemic from states

with the highest human incidence prior to June. We started the epidemic for the year 2015 by adding two infected nodes, one in sub-network *SN4* (California) and another in sub-network *SN42* (Texas). Connections among sub-networks are developed by distance dispersal kernels. Parameters for these kernels are estimated from the ABC-SMC method.

3.2 ABC-SMC for parameter estimation and model comparison

3.2.1 Parameter estimation

ABC-SMC parameter estimation was applied to three dispersal kernel network models separately. For each set of prior distributions, convergence to the posterior distribution was achieved after 13-15 SMC rounds. Convergence of the posterior distributions was monitored by visual inspection of the outputs from consecutive SMC rounds. The prior distribution for exponential network parameter was, $K_e \sim U(0.1, 0.3)$, for power-law $K_{pl} \sim U(2, 4)$, for power-law biased by flyway was $K_{pl} \sim U(2, 4)$. Prior distribution for constant of transmission rate β_0, transition rate from exposed to infectious λ, and human spillover rate η is same for three kernel models; $\beta_0 \sim U(0, 15)$, $\lambda \sim U(0.025, 10)$ and $\eta \sim U(0, 50)$. Perturbation kernels were also uniform, $PK = \alpha U(-1, 1)$, with $\alpha = 0.5(max\theta_{p-1} - min\theta_{p-1})$, here θ_{p-1} is the set of a parameter values in the previous population. We used weekly human case data for 49 locations, as observed data. The estimated parameters for this three dispersal kernel network models for 2015 are presented in Table 3.1.

3.2.2 Model comparison

ABC-SMC for model selection allows us to estimate posterior model distributions. We used this algorithm to compare the three distance kernels. Prior distributions and perturbation kernels are the same for both the model selection and the parameter estimation algorithm. Here we used one more prior distribution for discrete model parame-

Table 3.1: **Estimated parameters for the year 2015 from ABC-SMC parameter estimation.** *Estimated using data from the Centers for Disease Control and Prevention (CDC),[1] the National Centers for Environmental Information,[2] and Clements et al..[3]

Parameter	Exponential	Power-law	Power-law biased by flyway	Source
Network Parameter, K				
mean	0.1264	3.3844	2.3147	
median	0.1216	3.3924	2.2690	Estimated*
(95% CI)	(0.1235, 0.1294)	(3.3329, 3.4260)	(2.3030, 2.3264)	
Constant for transmission rate, β_0				
mean	0.0439 day^{-1}	0.2026 day^{-1}	0.0059 day^{-1}	
median	0.0362 day^{-1}	0.0526 day^{-1}	0.0061 day^{-1}	Estimated*
(95% CI)	(0.0354, 0.0524 day^{-1})	(0.0574, 0.3478 day^{-1})	(0.0058, 0.0059 day^{-1})	
Transition rate from exposed to infectious node, λ				
mean	0.0884 day^{-1}	0.1069 day^{-1}	0.0721 day^{-1}	
median	0.0823 day^{-1}	0.1059 day^{-1}	0.0706 day^{-1}	Estimated*
(95% CI)	(0.0820, 0.0948 day^{-1})	(0.0940, 0.1197 day^{-1})	(0.0718, 0.0724 day^{-1})	
Bird Recovery rate, δ				
range	0.2-0.25 day^{-1}	0.2-0.25 day^{-1}	0.2-0.25 day^{-1}	Komar et al.[55]
Human spillover, η				
mean	0.2175 day^{-1}	0.2141 day^{-1}	0.4558 day^{-1}	
median	0.2173 day^{-1}	0.2154 day^{-1}	0.4599 day^{-1}	Estimated*
(95% CI)	(0.2098, 0.2252 day^{-1})	(0.2071, 0.2210 day^{-1})	(0.4479, 0.4637 day^{-1})	

ter; $m \sim U(1,3)$. The tolerance vector for ABC-SMC model selection algorithm is, $\epsilon = \{2200, 2000, 1800, 1600, 1400, 1200, 1100, 1000\}$. The target and intermediate distributions of model parameters are shown in Fig. 3.1.

We calculated the Bayes factor from the marginal posterior distribution of m, which we took from the final or last population. In the final population for 2015, exponential distance kernel model ($m = 1$) was selected for 64 times, power-law distance kernel ($m = 2$) was selected for 95 times and power-law influenced by flyway distance kernel model ($m = 3$) was selected for 841 times. Bayes factor $B_{3,1} = 841/64 = 13.1406, B_{3,2} = 841/95 = 8.8526$. In the marginal posterior distribution of three models, there is positive evidence in favor of power-law influenced by flyway distance kernel when compared with other two models.[15]

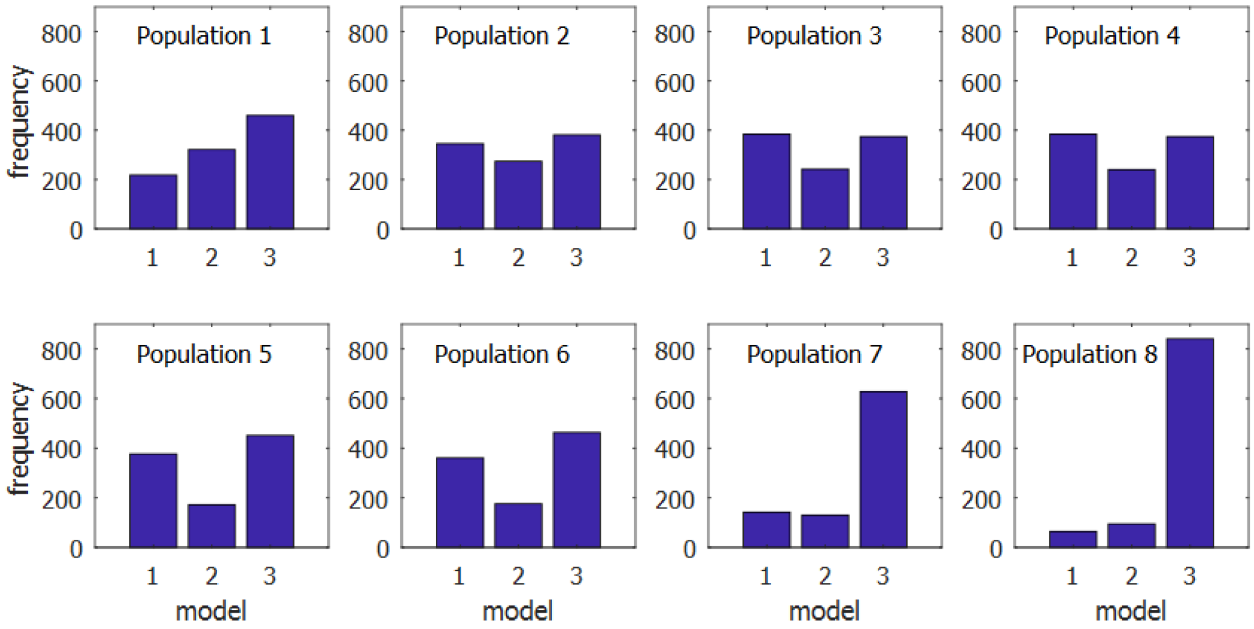

Figure 3.1: **Population of the marginal posterior distribution of the three models for the year 2015.** Model-1 represents exponential kernel, model-2 represents power-law kernel, and model-3 represents power-law influenced by flyway kernel. Here, Population-8 is the approximation of the final marginal posterior distribution of model parameter m and population 1-7 are intermediate distributions. Population-0 is the discrete uniform prior distribution, which is not shown here.

The distribution of parameters for power-law influenced by flyway for 2015 are presented in Fig. 3.2. Calculation of the Bayes factor for 2014 and 2016 are provided in the Appendix A.

3.3 Performance of the power-law-flyway network model

To test the performance of this framework, we used estimated parameters from Table A.1 for power-law kernel influenced by flyway. We set the parameters value; $K_{pl} = 2.3147$, $\beta_0 = 0.0059 day^{-1}$, $\lambda = 0.0721 day^{-1}$, and $\delta = 0.2031 day^{-1}$. The simulation period for the avian population model is from week-23 to week-44. The output of avian population was used as the input of zoonotic spillover compartment. Then we compared the output of zoonotic spillover compartment with human case data for week 24 to week 45. We considered a one-week lag between WNV incidence in birds and WNV incidence in humans. In humans, WNV-infected individuals (approximately 20%) develop a mild febrile illness after

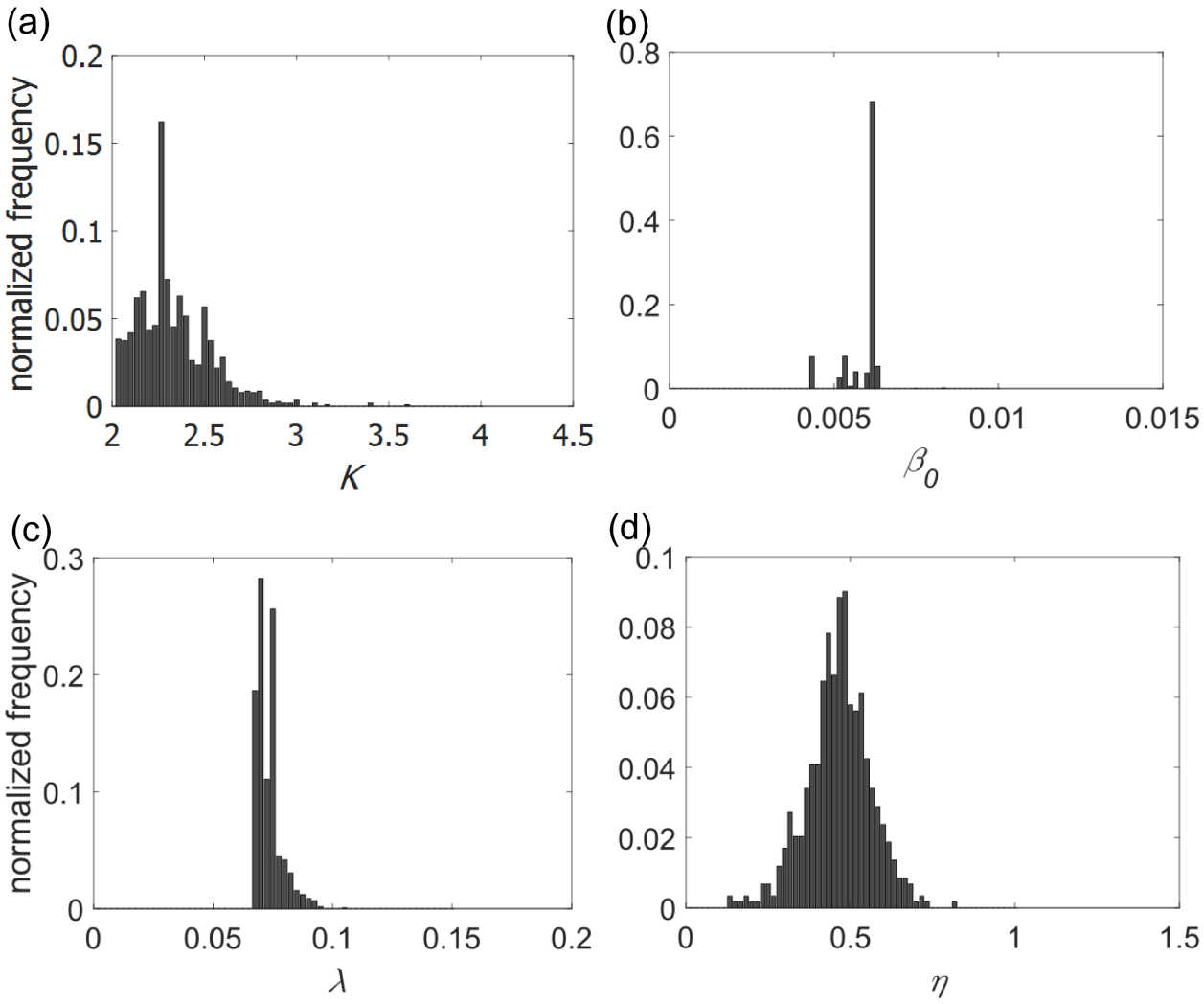

Figure 3.2: **Histograms of the approximated posteriors distribution of parameters for power-law influenced by flyway kernel for the year 2015.** a) Network Parameter K ; b) constant for transmission rate β_0; c) transition rate from exposed to infectious node λ, and d) human spillover η.

36 days.[56] Peak of reporting of dead birds is one week prior than the reporting peak of human incidence.[57] In Fig. 3.3, the mean simulated human case from the 49 sub-networks is compared with the weekly human case data for 2015 for the contiguous USA. The absolute errors between them are shown here. From this whisker plot, we can see that the median of the absolute error for the states is close to zero. In Fig. 3.3, the largest outlier is California (marked by black circles). These outliers result from a mismatch between the simulated peak human incidence time and the observed human incidence peak time possibly because

the very long state (north to south) has weather which is very different in southern California (warmer and drier) than northern California (cooler and wetter) causing a difference between peak mosquito seasons in the southern and northern parts.

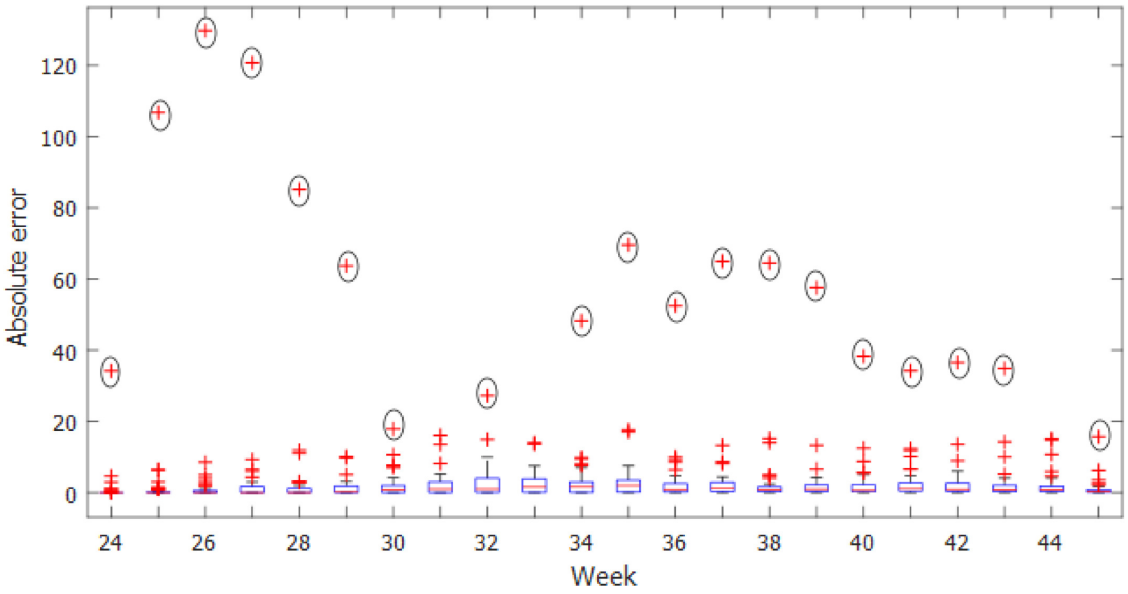

Figure 3.3: **Absolute errors of the simulated human cases of 49 states by weeks with the observed data for the year 2015.** Mean of 1000 realizations has used as the simulated data. On the blue boxes, the red horizontal lines show the median and the bottom and top edges of the boxes indicate 25^{th} and 75^{th} percentile respectively. The whiskers show the ranges of data points not considered outliers and outliers are showing by red + symbol. Californian outliers are marked by black circles.

We compared the total yearly incidence of human WNV from this model with the state level reported case data. The results are shown in Fig. 3.4. For 2015, we found that the case data for 42 of 49 locations were within the simulation results. The states where human cases were different from the simulation results were *over-reported* states (Nevada) and *under-reported* states (Louisiana, Mississippi, Nebraska, North Dakota, South Dakota, and Washington). The possible reason for this mismatch are reporting error or overwintering of virus in birds or mosquitoes or another bird species (not robins) is the key reservoir species for that state

To build a disease prevalence map, we grouped the states in four categories; 1) higher

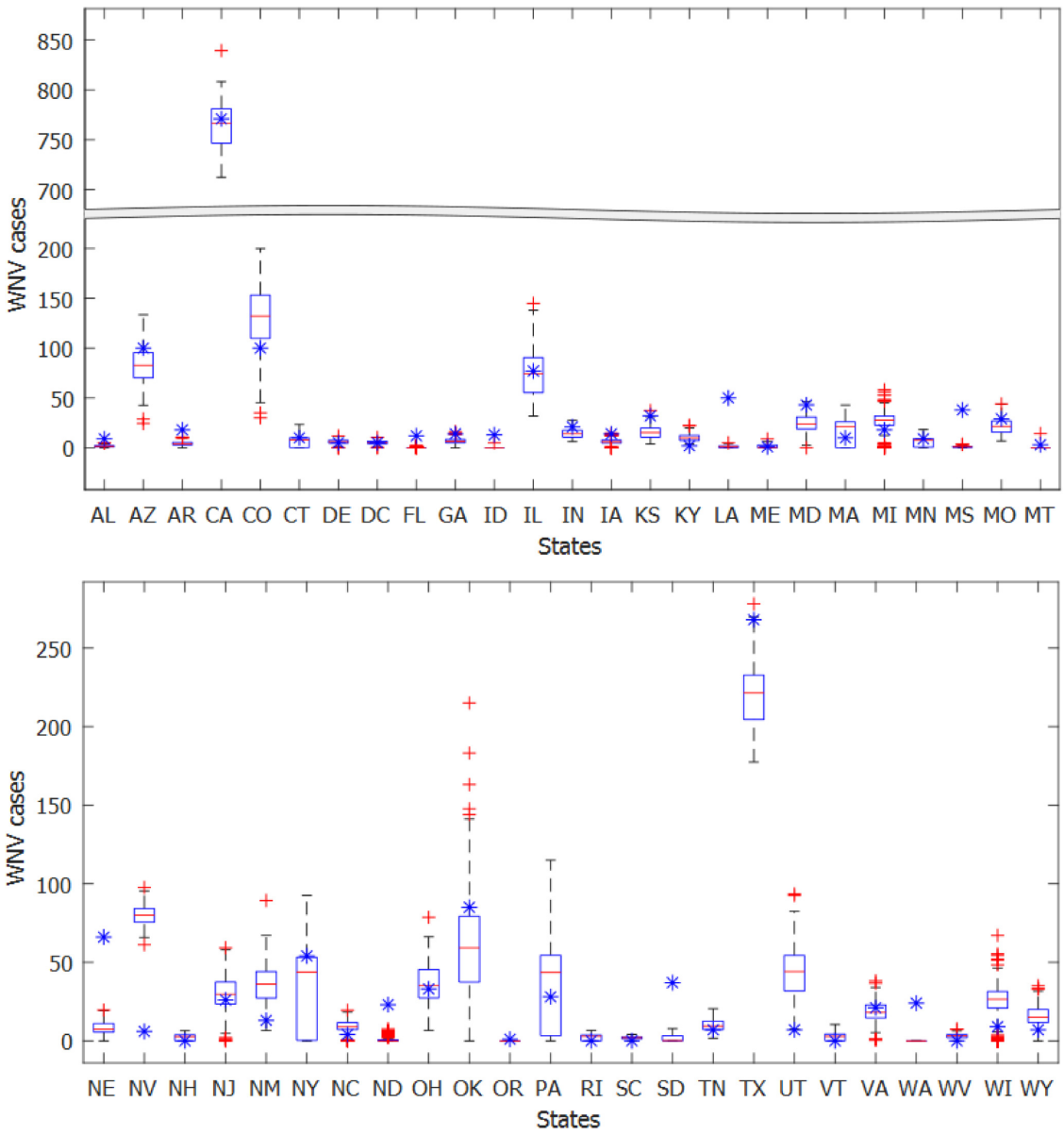

Figure 3.4: **WNV human incidence by states for the year 2015 from power-law influenced by flyway kernel model** (for K_{pl}=2.3147, $\beta_0 = 0.0059day^{-1}$, $\lambda = 0.0721day^{-1}$, $\eta = 0.4558day^{-1}$), generated from 1000 simulation and observed data are indicated by blue colored star points. states name are given in the short form. Simulated results are represented with a box plot in which the red horizontal lines show the median and the bottom and top edges of the boxes indicate 25^{th} and 75^{th} percentile respectively, The whiskers show the ranges of data points not considered outliers and outliers are showing by red + symbol. Broken scale is used for sake of visualization.

prevalence —incidence is more than 100, 2) intermediate prevalence —incidence is in between 50-99, 3) moderate prevalence —incidence is in between 25-49 and 4) low prevalence —

incidence is less than 25. To group the states, we used the median of the simulation results. The disease prevalence map from the model are presented in Fig. 3.5a and from observed data are presented in Fig. 3.5b. Among 49 locations, 40 locations are in the same prevalence group in both maps.

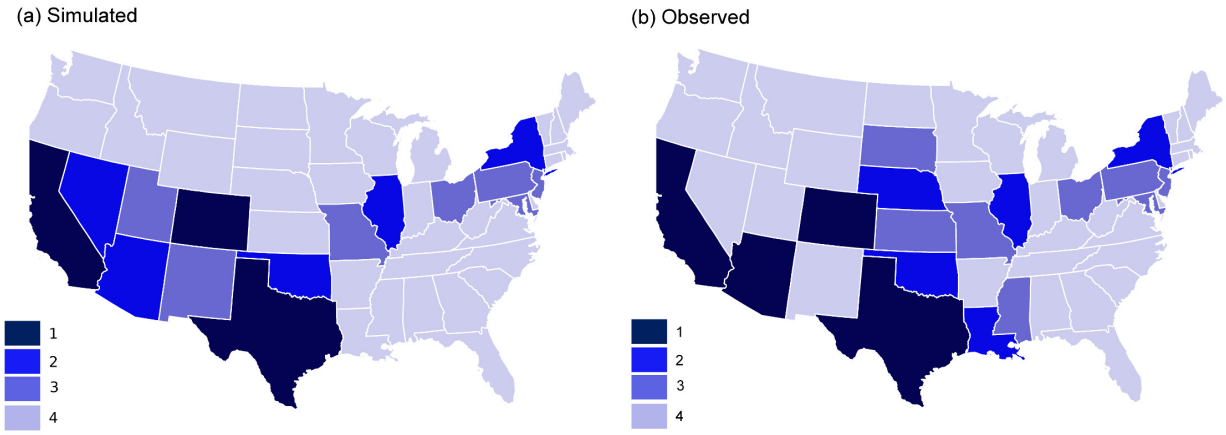

Figure 3.5: **Disease prevalence map for WNV human incidence for the year 2015.** The darker regions imposed greater prevalence. States are divided into four groups by incidence number; group-1: more than 99, group-2: 50-99. group-3: 25-49, and group-4: less than 25 incidences. a) States are divided by the median of the output of 1000 simulations, b) states are divided by observed data.

3.4 Mitigation strategies

We applied mitigation strategies on the power-law-flyway kernel network model to find the optimal mitigation plan. Fig. 3.6a shows the number of infected states or epidemic size for dynamic infected places tracing. Epidemic size decreased faster with increased reduction factor for *case-2*(infected & first neighbors) and *case-3*(infected & first neighbors & second neighbors) than *case-1*(only infected). The number of states where control measures were applied is displayed in Fig. 3.7, which is proportional to cost. Therefore, the cost was minimal for case-2 than other two cases for $RF > 2$. From the cost analysis, we concluded that, although the cost for *case-1* is less at the beginning of the yearly outbreak, we need to apply management only in the infected places, however by the end of the year the total cost for *case-2* will smaller because of the smaller epidemic size.

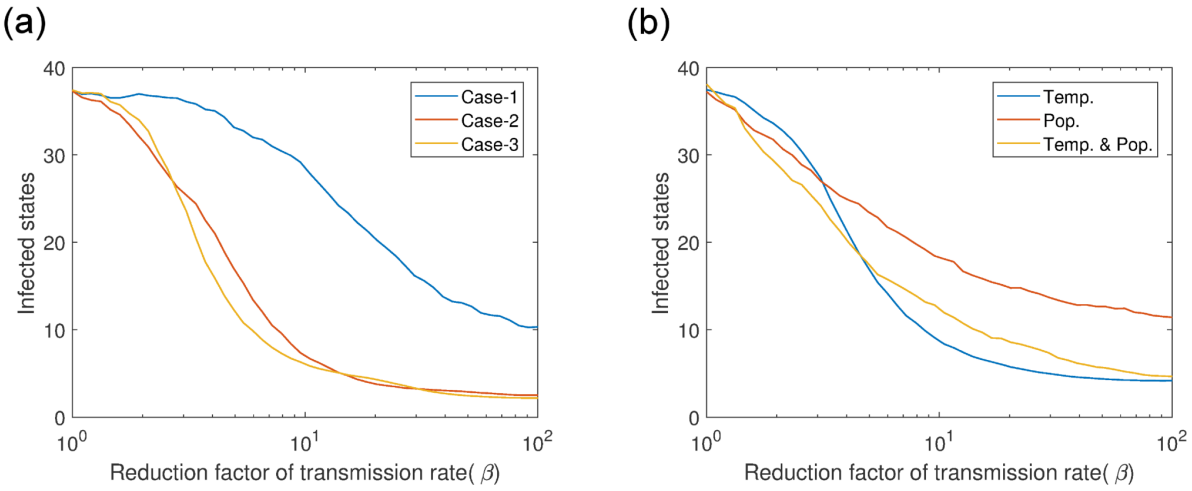

Figure 3.6: **Infected states for two types of mitigation strategies;** a) Dynamic infected places tracing; case-1: control measures are applied only in the infected states, case-2: control measures are applied in the infected states plus in their first neighboring states, case-3: control measures are applied in the infected places plus in their first and second neighboring states, and b) static ranked based strategy –states are ranked by; 1) temperature (Temp.), 2) avian population size (Pop.), 3) both(Temp & Pop.), then control measures are applied in the top 30% states. Log scale has used in x-axis for better visualization.

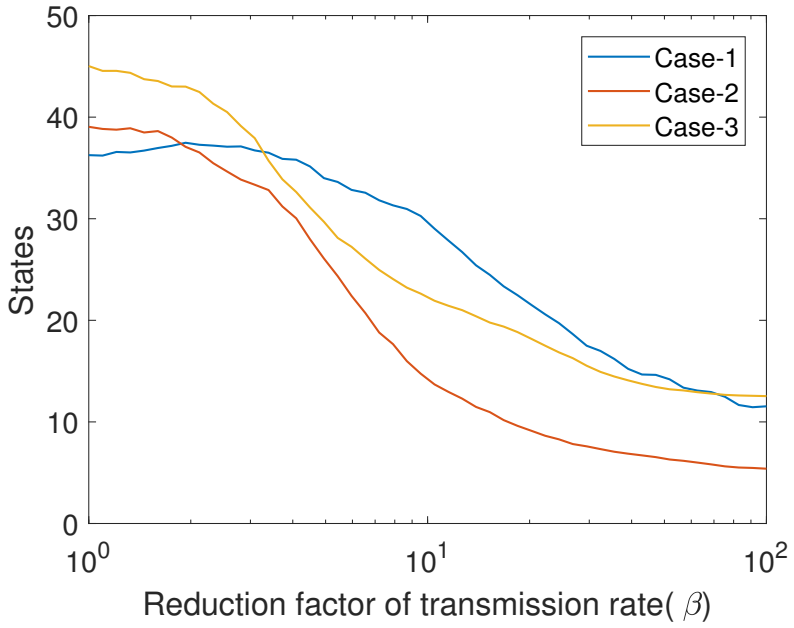

Figure 3.7: **Number of states where control measures are applied for the infected places tracing mitigation strategy.** Log scale has used in x-axis for better visualization.

The results of the static ranked based mitigation strategy measure are presented in Fig. 3.6b. We observed that, before $RF = 4.5$, number of infected states for *temp. & pop.*

dropped earlier than others. Number of infected states or epidemic size was smaller for *temp.* than *pop.* after $RF > 3$, infected population of a sub-network are more positively correlated with temperature. The NS_c is always the same for these three cases. For all mitigation strategies, minimum epidemic size could be 2, as we started the epidemic from two states.

Chapter 4

Discussion

We proposed an individual-based heterogeneous network framework and tested three dispersal kernels to understand the spatial spread patterns of WNV human case data across the contiguous United States.

This framework requires fewer parameters and has more flexibility to represent the spatial-temporal dynamics of WNV. Adding parameters will make the framework more realistic, for example, more competent bird species, landscape features for habitat preferences of host and vector species, daylight conditions,[32] pathogen invasion from outside of USA, variable susceptibility among different hosts and vectors, WNV strain variability, mosquito and virus overwintering, vertical transmission, human movement characteristics etc.. However, inclusion of too many factors increases model complexity which makes model optimization difficult given the availability of limited observational data. On the other hand, a simple model may insufficient to represent WNV spatial dynamics. Computational models need to be developed and parameters calculated with sufficient detail to be biologically accurate if they are used to evaluate epidemic management measures. However, for most biological systems, reliable parameter information is unknown. Unknown parameters or inaccurate assumptions add uncertainty to the model. Our framework has only four parameters to estimate (network Parameter K, transmission rate β, transition rate from exposed to infec-

tious state, λ, and human spillover, η). This framework has compartments only for the avian population (susceptible, exposed, infected, and recovered), which does not have to be species specific. We reduced the compartments for vector population by implementing them implicitly through transmission rate between infected nodes and susceptible nodes. The presented framework and dispersal kernel network model has an intermediate complexity that approximate Bayesian computation based on sequential Monte Carlo sampling (ABC-SMC) method successfully calibrated and estimated the parameters with the available data. If more data becomes available, it is possible to add them in this model for improved performance of the model.

Furthermore, this framework is flexible and therefore can represent various hosts and vectors including with population seasonality, which plays an important role in WNV dynamics. For host population seasonality, we added a node property *Activity*, this property allows us to control active host populations in the network in a specific time period. We added vector seasonality in this framework through temperature dependent transmission rate. This framework proposed one exponential and two fat-tailed distance kernel models for long-distance transmission of WNV. WNV spatial distribution is very complex because WNV can infect more than 300 bird species, some of which are residential birds and short-distance migrators which disperse less than 500 km distances (short connections) whereas some species are long-distance migratory birds creating long connections. The long-distance migratory birds are the long-distance dispersal (LDD) agents for WNV. Previous studies tried to analyze spreading of WNV using a traveling wave with constant velocity, however, WNV spread more rapidly across the North America than would be expected from the assumption of constant velocity traveling wave.[58] Likely this is because traveling wave models unlike distance dispersal kernel models for WNV spreading do not capture the long-distance migrating birds which can have various migratory ranges and distances. Distance dispersal kernels have more flexibility to represent the different bird migration distances and can account for accelerating invasions. However, exponential kernels produce short-connections and therefore like traveling waves are limited to constant expansion, unlike fat-tailed power-law kernels which can generate accelerating invasions by creating the long-distance connections from migratory

birds.[59] However, a general fat-tailed power-law kernel makes long-distance links in every direction which does not follow the incidence of WNV. Instead, a power-law-flyway kernel can be used to produce the long connections in the direction of flyways and short links in other directions. Bayesian inference was used to test which of the three kernel models best described WNV distribution on the network for three most recent years (2014- 2016). The power-law-flyway kernel best described the distribution of WNV cases because the long-range WNV transmission was concentrated mainly along the migratory bird flyways. The general power-law kernel overestimated the incidence data in some states because it was creating long-distance links in all directions.

The performance for the power-law-flyway dispersal kernel model was evaluated for the three most recent years (2014-2016) when WNV was endemic in the USA. The observed case data for the 49 locations were within the range of the simulated results for 41 states for 2014 (Fig A.2), 42 states for 2015 (Fig. 3.4), and 45 states for 2016 (Fig A.4). For all three years, the simulated results were similar to the observed data, except in Colorado, Louisiana, Mississippi, Nevada, Nebraska, North Dakota, and Washington. Nevada was *over-reported* for 2015 and all others were *under-reported*. The power law flyway dispersal kernel network model reported more WNV human incidence in Nevada than reported cases, one possible reason for over-reporting cases in Nevada has rural areas, which tend to under report human cases, whereas mosquito control districts and health departments, focused in urban areas, must test birds and mosquitoes, which explains why CDC reported WNV infected mosquitoes in 25% of counties in Nevada. The under-reported states had more human cases than predicted by the model. Under-reporting by the power-law-flyway kernel network model is likely because overwintering of the virus in some states (for example, Louisiana, Mississippi etc.), which was not considered. The overwintering infected *Culex* mosquitoes can stay in hibernacula such as sewers, houses, caves, and other warm areas in urban, suburban, and rural areas and initiate the outbreak in the spring. Furthermore, there may be under-reporting of cases by the model if robins are not the main reservoir species in a state, which would be predicted between gulf coast states (Louisiana and Mississippi) and northern

states such as North and South Dakota and Washington.

Mitigation strategies for WNV were tested using the power-law dispersal kernel network model. The management measures are not specific to larvae or adults, rather simply generally accepted best practices to reduce mosquito abundance for the purpose of reducing pathogen transmission. The mitigation strategy analysis proposes supplemental measures in addition to the existing mosquito management in each state because the states had yearly reported WNV cases despite the existing management methods. To reduce WNV spread, a theoretical policy would be management in neighboring regions and not exclusively in the infected places. Although this approach can cost more at the beginning of the epidemic season however at the end, it can reduce total cost by decreasing the size of the epidemic. If management measures are applied only in the infected states, it is not possible to control the epidemic because of long-distance migratory birds. This is a statewide management in a unified effort. We acknowledge that states do not conduct mosquito management in this way, but to test the spillover it was necessary to do the simulation in this way because only state level data was available.

Cooperation and communication equal early treatment and reduced outbreak sizes because of reduced WNV dispersal by American robins. This novel model can be applied to find out the invasion patterns of other long-distance dispersing pathogens.

Bibliography

[1] Centers for disease control and prevention. URL https://www.cdc.gov/westnile/index.html. Accessed: 2017-09-30.

[2] National centers for environmental information. URL https://www.ncdc.noaa.gov/. Accessed: 2017-05-13.

[3] JF Clements, TS Schulenberg, MJ Iliff, D Roberson, TA Fredericks, BL Sullivan, and CL Wood. The ebird/clements checklist of birds of the world: v2015. *URL: http://www. birds. cornell. edu/clementschecklist/download/IOC*, 2015.

[4] Alexis Burakoff. West nile virus and other nationally notifiable arboviral diseasesunited states, 2016. *MMWR. Morbidity and Mortality Weekly Report*, 67, 2018.

[5] Louis D Bergsman, James M Hyman, and Carrie A Manore. A mathematical model for the spread of west nile virus in migratory and resident birds. *Math Biosci Eng*, 13(2): 401–24, 2016.

[6] Nicholas Komar, Stanley Langevin, Steven Hinten, Nicole Nemeth, Eric Edwards, Danielle Hettler, Brent Davis, Richard Bowen, and Michel Bunning. Experimental infection of north american birds with the new york 1999 strain of west nile virus. *Emerging infectious diseases*, 9(3):311, 2003.

[7] Michael J Turell, David J Dohm, Michael R Sardelis, Monica L Oguinn, Theodore G Andreadis, and Jamie A Blow. An update on the potential of north american mosquitoes (diptera: Culicidae) to transmit west nile virus. *Journal of medical entomology*, 42(1): 57–62, 2005.

[8] Georg Pauli, Ursula Bauerfeind, Johannes Blümel, Reinhard Burger, Christian Drosten,

Albrecht Gröner, Lutz Gürtler, Margarethe Heiden, Martin Hildebrandt, Bernd Jansen, et al. West nile virus. *Transfusion medicine and hemotherapy*, 40(4):265, 2013.

[9] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.

[10] Marjorie J Wonham, Tomas de Camino-Beck, and Mark A Lewis. An epidemiological model for west nile virus: invasion analysis and control applications. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(1538):501–507, 2004.

[11] Rongsong Liu, Jiangping Shuai, Jianhong Wu, and Huaiping Zhu. Modeling spatial spread of west nile virus and impact of directional dispersal of birds. *Mathematical Biosciences and Engineering*, 3(1):145, 2006.

[12] Mark Lewis, Joanna Rencławowicz, and P Van den Driessche. Traveling waves and spread rates for a west nile virus model. *Bulletin of mathematical biology*, 68(1):3–23, 2006.

[13] Benoit Durand, Gilles Balança, Thierry Baldet, and Véronique Chevalier. A metapopulation model to simulate west nile virus circulation in western africa, southern europe and the mediterranean basin. *Veterinary research*, 41(3):32, 2010.

[14] Norberto Aníbal Maidana and Hyun Mo Yang. Spatial spreading of west nile virus described by traveling waves. *Journal of theoretical biology*, 258(3):403–417, 2009.

[15] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.

[16] Katalin Csilléry, Michael GB Blum, Oscar E Gaggiotti, and Olivier François. Approximate bayesian computation (abc) in practice. *Trends in ecology & evolution*, 25(7): 410–418, 2010.

[17] Mark A Beaumont. Approximate bayesian computation in evolution and ecology. *Annual review of ecology, evolution, and systematics*, 41:379–406, 2010.

[18] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate bayesian computation. *PLoS computational biology*, 9(1):e1002803, 2013.

[19] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[20] Scott A Sisson, Yanan Fan, and Mark M Tanaka. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.

[21] Charles S Apperson, Hassan K Hassan, Bruce A Harrison, Harry M Savage, Stephen E Aspen, Ary Farajollahi, Wayne Crans, Thomas J Daniels, Richard C Falco, Mark Benedict, et al. Host feeding patterns of established and potential mosquito vectors of west nile virus in the eastern united states. *Vector-Borne and Zoonotic Diseases*, 4(1):71–82, 2004.

[22] A Marm Kilpatrick, Peter Daszak, Matthew J Jones, Peter P Marra, and Laura D Kramer. Host heterogeneity dominates west nile virus transmission. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1599):2327–2333, 2006.

[23] Harry M Savage, Deepak Aggarwal, Charles S Apperson, Charles R Katholi, Emily Gordon, Hassan K Hassan, Michael Anderson, Dawn Charnetzky, Larry McMillen, Emily A Unnasch, et al. Host choice and west nile virus infection rates in blood-fed mosquitoes, including members of the culex pipiens complex, from memphis and shelby county, tennessee, 2002–2003. *Vector-Borne and Zoonotic Diseases*, 7(3):365–386, 2007.

[24] Gabriel L Hamer, Uriel D Kitron, Tony L Goldberg, Jeffrey D Brawn, Scott R Loss, Marilyn O Ruiz, Daniel B Hayes, and Edward D Walker. Host selection by culex pipiens

mosquitoes and west nile virus amplification. *The American journal of tropical medicine and hygiene*, 80(2):268–278, 2009.

[25] Gustavo Cruz-Pacheco, Lourdes Esteva, Juan Antonio Montaõ-Hirose, and Cristobal Vargas. Modelling the dynamics of west nile virus. *Bulletin of mathematical biology*, 67 (6):1157, 2005.

[26] Nicholas B DeFelice, Eliza Little, Scott R Campbell, and Jeffrey Shaman. Ensemble forecast of human west nile virus cases and mosquito infection rates. *Nature Communications*, 8, 2017.

[27] Michael R Sardelis, Michael J Turell, David J Dohm, and Monica L O'Guinn. Vector competence of selected north american culex and coquillettidia mosquitoes for west nile virus. *Emerging infectious diseases*, 7(6):1018, 2001.

[28] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

[29] Caterina M Scoglio, Claudio Bosca, Mahbubul H Riad, Faryad D Sahneh, Seth C Britch, Lee W Cohnstaedt, and Kenneth J Linthicum. Biologically informed individual-based network model for rift valley fever in the us and evaluation of mitigation strategies. *PloS one*, 11(9):e0162759, 2016.

[30] Mahbubul H Riad, Caterina M Scoglio, D Scott McVey, and Lee W Cohnstaedt. An individual-level network model for a hypothetical outbreak of japanese encephalitis in the usa. *Stochastic environmental research and risk assessment*, 31(2):353–367, 2017.

[31] Peter M Rabinowitz, Deron Galusha, Sally Vegso, Jennifer Michalove, Seppo Rinne, Matthew Scotch, and Michael Kane. Comparison of human and animal surveillance data for h5n1 influenza a in egypt 2006–2011. *PloS one*, 7(9):e43851, 2012.

[32] Hamid Reza Nasrinpour, Alexander A Reimer, Marcia R Friesen, and Robert D McLeod. Data preparation for west nile virus agent-based modelling: Protocol for processing bird

population estimates and incorporating arcmap in anylogic. *JMIR research protocols*, 6 (7), 2017.

[33] Marcus SC Blagrove, Cyril Caminade, Elisabeth Waldmann, Elizabeth R Sutton, Maya Wardeh, and Matthew Baylis. Co-occurrence of viruses and mosquitoes at the vectors optimal climate range: An underestimated risk to temperate regions? *PLoS neglected tropical diseases*, 11(6):e0005604, 2017.

[34] LM Rueda, KJ Patel, RC Axtell, and RE Stinner. Temperature-dependent development and survival rates of culex quinquefasciatus and aedes aegypti (diptera: Culicidae). *Journal of medical entomology*, 27(5):892–898, 1990.

[35] William K Reisen, Ying Fang, and Vincent M Martinez. Effects of temperature on the transmission of west nile virus by culex tarsalis (diptera: Culicidae). *Journal of medical entomology*, 43(2):309–317, 2006.

[36] William K Reisen. Effect of temperature on culex tarsalis (diptera: Culicidae) from the coachella and san joaquin valleys of california. *Journal of medical entomology*, 32(5): 636–645, 1995.

[37] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[38] Albert-László Barabási. Network science book. *Network Science*, 625, 2014.

[39] Samuel Soubeyrand, Leonhard Held, Michael Höhle, and Ivan Sache. Modelling the spread in space and time of an airborne plant disease. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(3):253–272, 2008.

[40] Irina Chis Ster and Neil M Ferguson. Transmission parameters of the 2001 foot and mouth epidemic in great britain. *PLoS One*, 2(6):e502, 2007.

[41] Sebastian Meyer, Leonhard Held, et al. Power-law models for infectious disease spread. *The Annals of Applied Statistics*, 8(3):1612–1639, 2014.

[42] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[43] Tommy Tsan-Yuk Lam, Hon S Ip, Elodie Ghedin, David E Wentworth, Rebecca A Halpin, Timothy B Stockwell, David J Spiro, Robert J Dusek, James B Bortner, Jenny Hoskins, et al. Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among north american birds. *Ecology letters*, 15(1):24–33, 2012.

[44] Mathieu Fourment, Aaron E Darling, and Edward C Holmes. The impact of migratory flyways on the spread of avian influenza virus in north america. *BMC evolutionary biology*, 17(1):118, 2017.

[45] Frederick C Lincoln. *Migration of birds*. Number 16. Government Printing Office, 1999.

[46] Sotirios Tsiodras, Theodoros Kelesidis, Iosif Kelesidis, Ulf Bauchinger, and Matthew E Falagas. Human infections associated with wild birds. *Journal of Infection*, 56(2):83–98, 2008.

[47] John H Rappole and Z Hubalek. Migratory birds and west nile virus. *Journal of applied microbiology*, 94(s1):47–58, 2003.

[48] Ellen Brooks-Pollock, Gareth O Roberts, and Matt J Keeling. A dynamic model of bovine tuberculosis spread and control in great britain. *Nature*, 511(7508):228, 2014.

[49] Chris P Barnes, Daniel Silk, and Michael PH Stumpf. Bayesian design strategies for synthetic biology. *Interface focus*, 1(6):895–908, 2011.

[50] Tina Toni and Michael PH Stumpf. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, 26(1):104–110, 2009.

[51] Anis Ben Abdessalem, Nikolaos Dervilis, David Wagg, and Keith Worden. Model selection and parameter estimation in structural dynamics using approximate bayesian computation. *Mechanical Systems and Signal Processing*, 99:306–325, 2018.

[52] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[53] Xavier Didelot, Richard G Everitt, Adam M Johansen, Daniel J Lawson, et al. Likelihood-free estimation of model evidence. *Bayesian analysis*, 6(1):49–76, 2011.

[54] Jean-Michel Marin, Natesh S Pillai, Christian P Robert, and Judith Rousseau. Relevant statistics for bayesian model choice. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(5):833–859, 2014.

[55] Nicholas Komar. West nile virus: epidemiology and ecology in north america. *Advances in virus research*, 61:185–234, 2003.

[56] G Dauphin and S Zientara. West nile virus: recent trends in diagnosis and vaccine development. *Vaccine*, 25(30):5563–5576, 2007.

[57] Glen D Johnson, Millicent Eidson, Kathryn Schmit, April Ellis, and Martin Kulldorff. Geographic prediction of human onset of west nile virus using dead crow clusters: an evaluation of year 2002 data in new york state. *American Journal of Epidemiology*, 163 (2):171–180, 2005.

[58] Christopher C Mundt, Kathryn E Sackett, LaRae D Wallace, Christina Cowger, and Joseph P Dudley. Long-distance dispersal and accelerating waves of disease: empirical relationships. *The American Naturalist*, 173(4):456–466, 2009.

[59] Mark Kot, Mark A Lewis, and Pauline van den Driessche. Dispersal data and the spread of invading organisms. *Ecology*, 77(7):2027–2042, 1996.

Appendix A

Simulation results for 2014 and 2016

In this research, we compared three distance dispersal kernel to understand the spatial distribution of WNV. Distance dispersal kernels are: 1) Exponential kernel, 2) power-law kernel, and 3) power-law influenced by flyway kernel. We used this framework in the USA for 2014-2016. Approximate Bayesian computation based on sequential Monte Carlo sampling (ABC-SMC) was used for parameter estimation and for selection of the best kernel. The results for 2014 and 2016 are given below.

A.1 Results for 2014

In 2014, WNV infected human cases in the USA was 2205. All the states were infected except Alaska, North Carolina, and West Virginia. The inputs of this framework for 2014 are weekly human incidence data by states, avian population data by states, and average monthly temperature data by states for 2014. The target and intermediate distributions of model parameter m from ABC-SMC model selection method are given in Fig. A.1. Bayes factor was calculated from the last population (population-8 in Fig. A.1). In the last population, exponential kernel was selected for 79 times, general power-law kernel was selected for 96 times, and power-law-flyway was selected for 825 times.

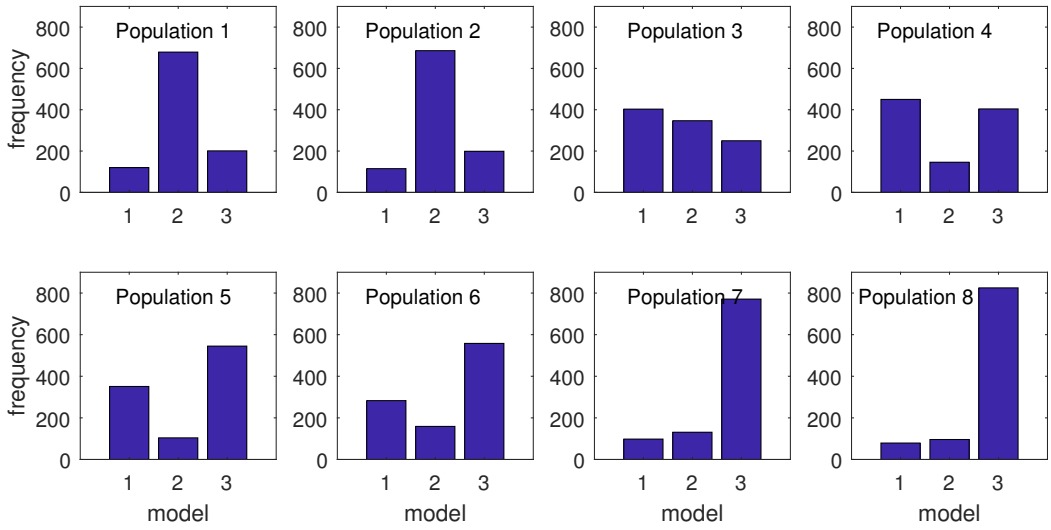

Figure A.1: **Population of the marginal posterior distribution of the three models for 2014.** Model-1 represents exponential kernel, model-2 represents power-law kernel, and model-3 represents power-law influenced by flyway kernel. Here, Population-8 is the approximation of the final marginal posterior distribution of model parameter m and population 1-7 are intermediate distributions. Population-0 is the discrete uniform prior distribution, which is not shown here.

The obtained Bayes factors for 2014 are:

$$B_{3,1} = \frac{825}{79} = 10.4430 \tag{A.1}$$

$$B_{3,2} = \frac{825}{96} = 8.5938 \tag{A.2}$$

From the analysis of Bayes factor for 2014, we can conclude that power-law influenced by flyway distance kernel network model has positive evidence against other two kernels. The estimated parameters are provided in Table A.1.

Performance of the power-law flyway kernel network model for 2014. To see the performance for this framework for 2014, we used estimated parameters from ABC-SMC parameter estimation method for power-law influenced by flyway kernel network model. The parameters are presented in the Table A.1. The results from 1000 simulation are aggregated in the box plot of Fig. A.2. Total observed human cases by states from CDC are given by blue

Table A.1: Estimated parameters for power-law biased by flyway network model for 2014, 2015 and 2016 from ABC-SMC parameter estimation algorithm.

Parameter		2014	2015	2016
Network parameter, K	mean	2.4091	2.3147	2.4233
	median	2.3495	2.2690	2.3889
	(95% CI)	(2.3469, 2.4713)	(2.3030, 2.3264)	(2.3353, 2.5114)
Constant for transmission rate, β_0	mean	0.0028 day^{-1}	0.0059 day^{-1}	0.0029 day^{-1}
	median	0.0026 day^{-1}	0.0061 day^{-1}	0.0031 day^{-1}
	(95% CI)	(0.0025, 0.0032 day^{-1})	(0.0058, 0.0059 day^{-1})	(0.0028, 0.0035 day^{-1})
Transition rate from exposed to infectious node, λ	mean	0.0445 day^{-1}	0.0721 day^{-1}	0.0452 day^{-1}
	median	0.0436 day^{-1}	0.0706 day^{-1}	0.0460day^{-1}
	(95% CI)	(0.0434, 0.0455 day^{-1})	(0.0718, 0.0724 day^{-1})	(0.0443, 0.0461 day^{-1})

star. We found that from 49 locations, the total human incidence case for 41 locations falls within the range of simulation results. The simulation results could not follow the observed data for Arizona, Colorado, Kansas, Louisiana, Mississippi, Nebraska, New Mexico, and Washington.

A.2 Results for 2016

In 2016, WNV infected human cases in the USA was 2149. All the states were infected for WNV (except Hawaii and Alaska). 27 states had more than 10 WNV disease cases. California, Colorado, Illinois, South Dakota, and Texas had more than 100 WNV disease cases. This is the most recent year when weekly WNV incidents are publicly available from *CDC*.[1] For host population, we used American Robin population data for 2016 from *eBird*.[3] The description of the host population and sub-networks are provided in Table S3 in the Text S3. We started the epidemic from Arizona because we found highest disease cases in Arizona among all other states before June (in this framework, simulation has started from

June). The target and intermediate distributions of model parameter m from ABC-SMC model selection method are shown in Supporting Fig. A.3. The Bayes factor is obtained from the marginal posterior distribution of m, which we got from the final population (Population 8 in Fig. A.3). The calculated Bayes factors are:

$$B_{3,1} = \frac{879}{88} = 9.9886 \tag{A.3}$$

$$B_{3,2} = \frac{879}{33} = 26.6364 \tag{A.4}$$

From the interpretation of Bayes factors,[52] we found positive evidence in favor of power-law influenced by flyway distance kernel network model compared to exponential distance kernel network model and strong evidence in favor of power-law influenced by flyway distance kernel network model compared to power-law distance kernel network model for 2016.

Performance of power-law flyway kernel network model for 2016. Fig. A.4 are showing the simulation results of 1000 realizations of the framework for 2016 for power-law-flyway network. We found that, observed data for 45 states among 49 locations falls within the range of the simulated results for 2016. The simulated results could not follow for observed human WNV incidence for Colorado, Louisiana, Mississippi, and North Dakota.

Table A.2: Summary of evidence among three network models from ABC-SMC model selection algorithm for 2014, 2015 and 2016.

Evidence	2014	2015	2016
Power-law-flyway kernel network model against exponential kernel network model	positive	positive	positive
Power-law-flyway kernel network model against power-law kernel network model	positive	positive	strong

A.3 Discussion

The results from ABC-SMC (approximate Bayesian computation with sequential Monte Carlo sampling) model selection method are similar for 2014, 1015, and 2016. The results from ABC-SMC model selection method are summarized in Table A.2. The ABC-SMC method selected power-law-flyway kernel as the best kernel than other two kernels. Power-law-flyway kernel can best describe the WNV human case data in the USA. The estimated parameters values from ABC-SMC parameter estimation method are slightly different for these three years. The reasons for this difference is the different initial condition, different host population, and seasonality (different temperature data).

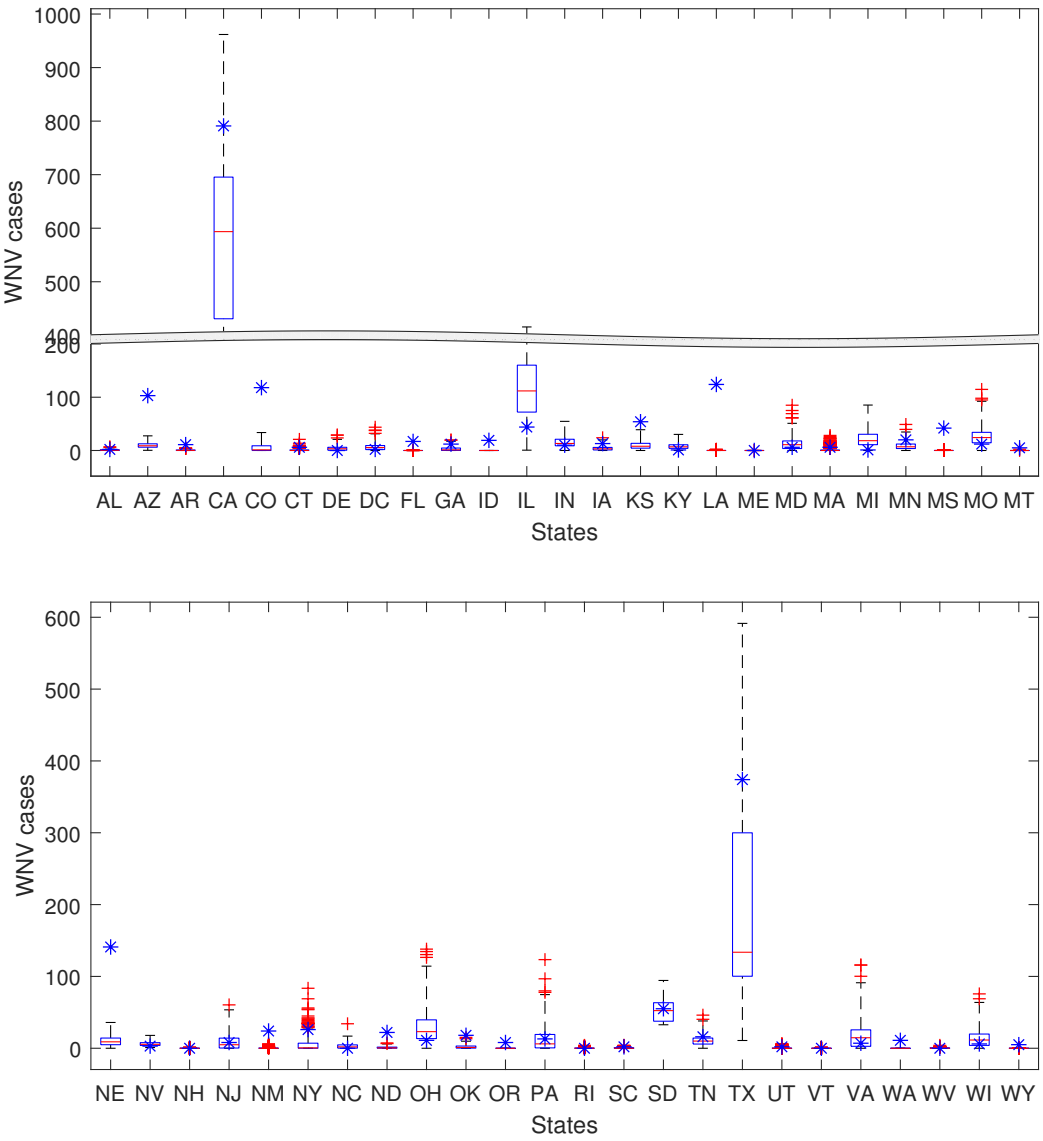

Figure A.2: **WNV human incidence by states for 2014 from power-law influenced by flyway kernel model** (for network parameter K=2.4091, constant for transmission rate $\beta_0 = 0.0028 day^{-1}$, and transition rate for exposed to infectious $\lambda = 0.0445 day^{-1}$), generated from 1000 simulation and observed data are indicated by blue colored star points. states name are given in short form. Simulated results are represented with a box plot in which the red horizontal lines show the median and the bottom and top edges of the boxes indicate 25^{th} and 75^{th} percentile respectively, The whiskers show the ranges of data points not considered outliers and outliers are showing by red + symbol. Broken scale is used for sake of visualization.

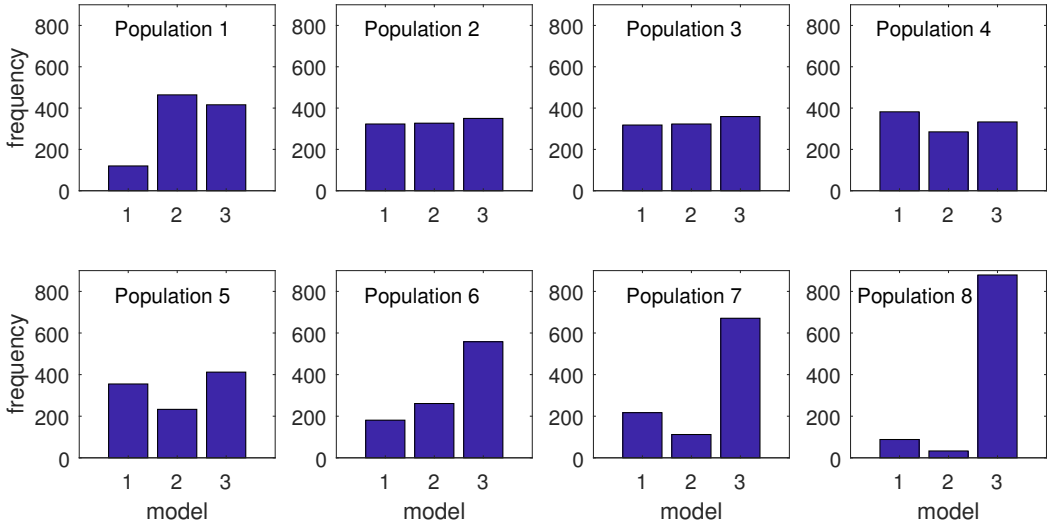

Figure A.3: **Population of the marginal posterior distribution of the three models for 2016.** Model-1 represents exponential kernel, model-2 represents power-law kernel, and model-3 represents power-law influenced by flyway kernel. Here, Population-8 is the approximation of the final marginal posterior distribution of model parameter m and population 1-7 are intermediate distributions. Population-0 is the discrete uniform prior distribution, which is not shown here.

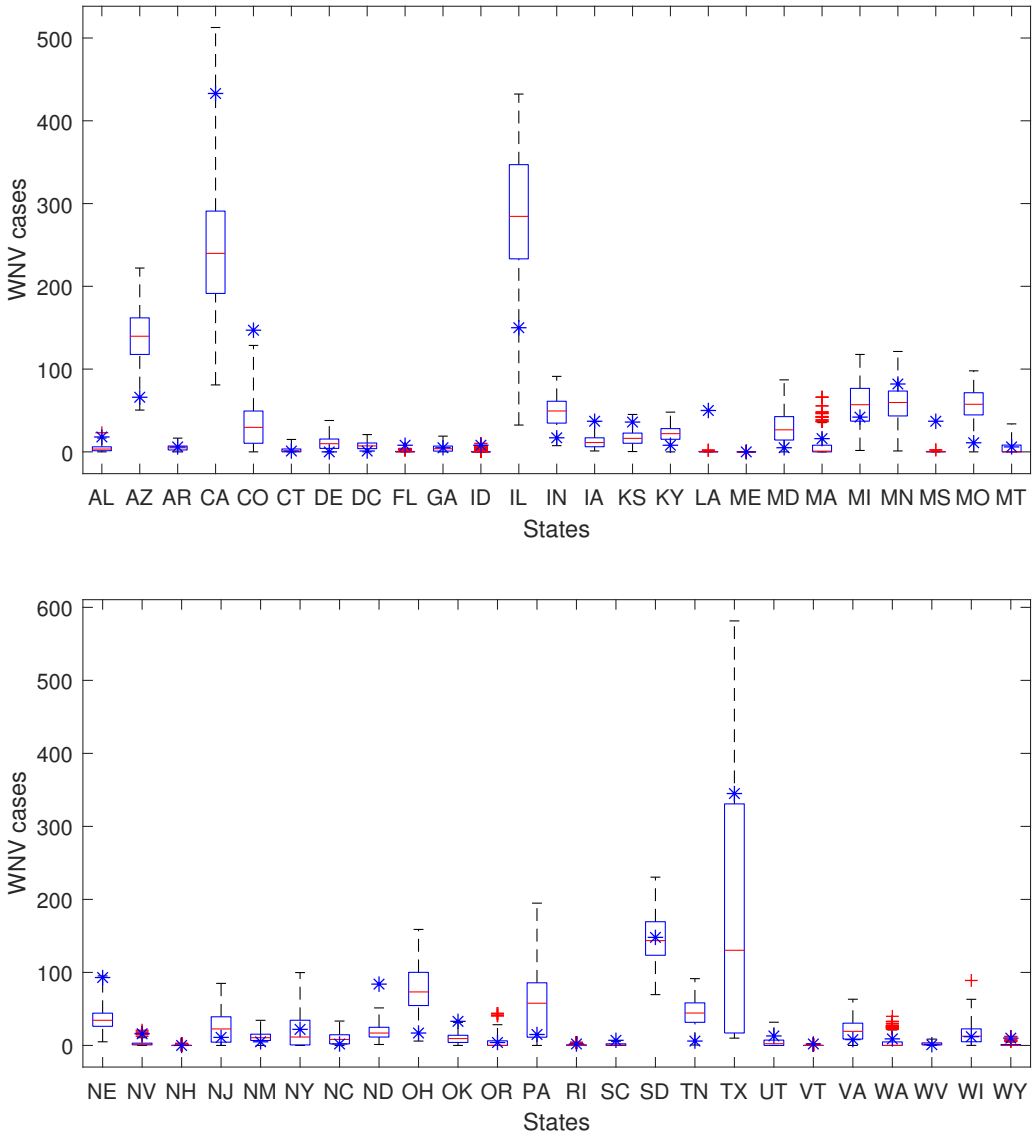

Figure A.4: **WNV human incidence by states for 2016 from power-law influenced by flyway kernel model** (for K=2.4233, $\beta_0 = 0.0029 day^{-1}$, and $\lambda = 0.0452 day^{-1}$), generated from 1000 simulation and observed data are indicated by blue colored star points. states name are given in short form. Simulated results are represented with a box plot in which the red horizontal lines show the median and the bottom and top edges of the boxes indicate 25^{th} and 75^{th} percentile respectively, The whiskers show the ranges of data points not considered outliers and outliers are showing by red + symbol.

Appendix B

Approximate Bayesian Computation based on sequential Monte Carlo sampling (ABC-SMC) method

This research has used ABC-SMC methods to estimate parameters and to select best kernel network model among three distance dispersal kernels.

B.1 Parameter estimation

The steps for approximate Bayesian computation with sequential Monte Carlo sampling (ABC-SMC) algorithm for parameter estimation are:[15–20]

Step 1 Initialize tolerance ϵ for each SMC round, where $\epsilon_P < ..... < \epsilon_2 < \epsilon_1$. Set Population indicator, p=0.

Step 2 Particle indicator, n=1.

Step 3 Generate a particle (set of parameters),θ_p^n

(a). if p=1, sample from prior of parameters, $\pi(\theta)$;

(b). if p>1, sample the particle from previous population $\{\theta^n_{p-1}\}$ with weights W_{p-1} and then perturb the particle, $\theta^{'}$ by using perturbation kernel, PK_p to get $\theta^{''}$.

(c). if $\pi(\theta^{''}) == 0$, return to Step 3.

Step 4 Run the model R times with the new particle and compare the simulated weekly human WNV incidence with observed weekly WNV incidence using the goodness-of-fit metric, We calculated $r_p(\theta^{''}) = (1/R) * \sum_{r=1}^{R} 1(dist(x,d) < \epsilon_p)$, if $r_p(\theta^{''}) == 0$ reject the particle; go back to Step 3(a).

Step 5 Calculate the weight for the accepted particle,

(a). if p=1, $W_{n,p} = r_p(\theta^{''})$;

(b). if p>1, the weight is given by, $W_{n,p} = \frac{\pi(\theta^i_p) * r_p(\theta^{''})}{\sum_{j=1}^{N} W_{i,p-1} PK_p(\theta^j_{p-1}, \theta^i_p)}$.

Step 6 Repeat steps 3 - 5 until N= 1000 particles have been accepted.

Step 7 Normalize the weights. If $p < P$, set p= p+1, go to Step 2.

B.2 Model seclection

The steps for approximate Bayesian computation with sequential Monte Carlo sampling (ABC-SMC) algorithm for model selection are:[15–20]

Step 1 Initialize tolerance for each SMC round $\epsilon_P < ..... < \epsilon_2 < \epsilon_1$. Set Population indicator, p=0.

Step 2 Particle indicator, n=1.

Step 3 Generate a particle

(a). if p=1, sample model parameter m and parameters for that model from prior, $\pi(m, \theta)$;

(b). if p>1, sample model $m^{'}$ with probability $Pr_{t-1}(m^{'})$ and then perturb by perturbation kernel PKm_p, sample the particle from previous population $\{\theta(m^{''})_{p-1}\}$ with weights W_{p-1} and then perturb the particle $\theta^{'}$ by using perturbation kernel to get $\theta^{''}$.

(c). if $\pi(m^{''}, \theta^{''}) == 0$, return to Step 3.

Step 4 Run the model $m^{''}$, R times with the new particle and compare the simulated weekly human WNV incidence with observed weekly WNV incidence using the goodness-of- fit metric, We calculated $r_p(\theta^{''}) = (1/R) * \sum_{r=1}^{R} 1(dist(x, d) < \epsilon_p)$, if $r_p(\theta^{''}) == 0$ reject the particle; go back to Step 3.

Step 5 Calculate weight for the accepted particle, set $(m_p^n, \theta_p^n) = (m^{''}, \theta^{''})$,

(a). if p=1, $W_{n,p}(m_p^n, \theta_p^n) = (1/R) * \sum_{r=1}^{R} 1(dist(x, d) < \epsilon_p)$; Here R is the number of replicate simulation run for a fixed particle.

(b). if p>1, the weight is given by, $W_{n,p}(m_p^n, \theta_p^n) = \frac{\pi(m_p^n, \theta_p^n) * (1/R) * \sum_{r=1}^{R} 1(dist(x,d) < \epsilon_p)}{\sum_{j=1}^{N} W_{i,p-1} PK_p(\theta_{p-1}^j, \theta_p^i)}$.

Step 6 Repeat steps 3 - 5 until N= 1000 particles have been accepted.

Step 7 Normalize the weights for every m. If $p < P$, set p= p+1, go to Step 3.

Appendix C

Network Description for 2014-2016

Table C.1: Description of sub-networks for 2014. $|V| = 9373$ and scaling constant $S_c = 0.03$ for 2014

Sub-network Id	Location	Nodes (% of $\|V\|$)	Probability of Active nodes				
			Jun.	Jul.	Aug.	Sept.	Oct.
SN1	AL	0.4588 %	0.4651	0.4884	0.4186	0.9767	1.0000
SN2	AZ	0.5334 %	1.0000	0.7600	0.4600	0.5200	0.5800
SN3	AR	0.3521 %	0.6970	1.0000	0.8182	0.6061	1.0000
SN4	CA	5.4732 %	0.6121	0.3996	0.3801	0.5166	1.0000
SN5	CO	2.2298 %	1.0000	1.0000	0.9139	1.0000	1.0000
SN6	CT	2.2511 %	1.0000	0.9716	0.7109	0.7630	1.0000
SN7	DE	1.4510 %	0.2868	0.3529	0.6397	1.0000	1.0000
SN8	DC	1.6964 %	0.2013	0.2830	0.6289	1.0000	0.8742
SN9	FL	0.1174 %	0.5455	0.4545	0.6364	0.6364	1.0000
SN10	GA	1.6003 %	0.5867	0.6600	0.5067	0.8400	1.0000
SN11	ID	0.6401 %	1.0000	1.0000	0.5333	1.0000	1.0000

SN12	IL	8.5778 %	0.7923	0.7923	0.8022	1.0000	1.0000
SN13	IN	1.7177 %	1.0000	0.9503	0.6584	0.9814	1.0000
SN14	IA	0.8855 %	0.8554	1.0000	0.7108	0.8313	1.0000
SN15	KS	1.2696 %	0.5378	0.7311	0.5882	1.0000	1.0000
SN16	KY	1.0135 %	0.9263	0.7158	0.8947	1.0000	1.0000
SN17	LA	0.2454 %	0.4783	0.4348	0.4348	0.5217	1.0000
SN18	ME	0.8322 %	1.0000	1.0000	0.5256	0.7308	1.0000
SN19	MD	3.3180 %	0.7878	0.8103	0.8842	1.0000	1.0000
SN20	MA	4.6623 %	0.8558	1.0000	0.9085	1.0000	1.0000
SN21	MI	4.0435 %	1.0000	0.9367	0.9604	1.0000	1.0000
SN22	MN	1.4403 %	1.0000	0.8296	0.6148	1.0000	1.0000
SN23	MS	0.1814 %	0.8235	0.5882	0.5882	0.6471	1.0000
SN24	MO	2.5605 %	0.5458	0.5417	0.4958	1.0000	1.0000
SN25	MT	1.3016 %	1.0000	1.0000	0.7131	0.6475	0.3689
SN26	NE	0.9175 %	1.0000	0.8837	0.6512	0.8488	1.0000
SN27	NV	0.3201 %	0.9667	0.7667	0.8000	1.0000	1.0000
SN28	NH	0.6828 %	1.0000	1.0000	0.6875	1.0000	1.0000
SN29	NJ	4.0435 %	0.6544	0.6781	0.7018	1.0000	1.0000
SN30	NM	0.5975 %	0.8036	0.6429	0.7857	1.0000	1.0000
SN31	NY	7.8843 %	0.8917	0.9107	1.0000	1.0000	1.0000
SN32	NC	1.4830 %	0.8633	0.7050	0.6835	1.0000	1.0000
SN33	ND	0.3627 %	1.0000	0.7647	1.0000	1.0000	1.0000
SN34	OH	4.4063 %	0.9637	1.0000	0.9274	1.0000	1.0000
SN35	OK	1.3016 %	0.2295	0.3033	0.3361	0.4098	1.0000
SN36	OR	2.0911 %	1.0000	1.0000	0.7041	1.0000	1.0000
SN37	PA	5.9853 %	0.8824	0.9037	1.0000	1.0000	1.0000
SN38	RI	0.4801 %	0.8000	1.0000	0.6444	0.8667	1.0000

SN39	SC	0.5334 %	0.4600	0.4600	0.5000	0.6600	1.0000
SN40	SD	0.4481 %	0.9762	1.0000	0.7619	0.9286	1.0000
SN41	TN	1.4403 %	0.6296	0.9259	0.8000	1.0000	1.0000
SN42	TX	1.6003 %	0.2333	0.3800	0.5333	0.6533	1.0000
SN43	UT	1.2056 %	1.0000	0.7345	0.3894	0.5044	0.8142
SN44	VT	0.8322 %	1.0000	1.0000	0.7436	0.8462	1.0000
SN45	VA	5.3558 %	0.3665	0.3546	0.5936	1.0000	1.0000
SN46	WA	3.8622 %	1.0000	1.0000	0.5884	1.0000	1.0000
SN47	WV	0.5868 %	1.0000	0.8000	0.4727	0.5455	1.0000
SN48	WI	4.2462 %	1.0000	0.7236	0.5829	0.9020	1.0000
SN49	WY	0.4801 %	1.0000	1.0000	0.8000	0.7778	0.3111

Table C.2: Description of sub-networks for 2015. $|V| = 7657$ and scaling constant $S_c = 0.02$ for 2015

Sub-network Id	Location	Nodes (% of $\|V\|$)	Probability of Active nodes				
			Jun.	Jul.	Aug.	Sept.	Oct.
SN1	AL	0.2612%	1.000	0.9000	0.9000	0.9956	1.000
SN2	AZ	0.6138%	1.000	0.9149	0.6595	0.7021	0.5106
SN3	AR	0.4832%	0.9730	0.6757	1.000	0.7027	1.000
SN4	CA	3.1213%	1.000	0.6192	0.6443	0.6066	1.000
SN5	CO	3.3172%	1.000	0.6181	0.4645	0.9094	0.7834
SN6	CT	2.1027%	1.000	0.8695	0.6894	0.6894	1.000
SN7	DE	0.8620%	1	0.6969	0.9242	1.0000	1.000
SN8	DC	0.7183%	0.5636	0.5636	1.0000	0.9989	1.000

SN9	FL	0.1437%	0.9091	1.0000	1.0000	0.9911	1.000
SN10	GA	1.0317%	1.0000	0.8101	0.7848	0.8354	1.000
SN11	ID	1.2146%	1.0000	0.7204	0.6344	1.0000	0.6666
SN12	IL	6.6083%	1.0000	0.7608	0.8023	1.0000	1.000
SN13	IN	1.3974%	1.0000	0.9345	0.8317	1.0000	1.000
SN14	IA	0.8880%	1.0000	0.6764	0.7794	1.0000	1.000
SN15	KS	0.9011%	1.0000	0.5942	0.8260	1.0000	1.000
SN16	KY	1.0709%	0.7073	0.7682	1.0000	1.0000	1.000
SN17	LA	0.2742%	0.6190	0.7619	1.0000	0.9978	1.000
SN18	ME	1.0709%	1.0000	0.8414	0.5609	0.4878	1.000
SN19	MD	3.3433%	0.6914	0.5898	1.0000	0.8906	1.000
SN20	MA	3.9963%	0.7778	0.8725	1.0000	1.0000	1.000
SN21	MI	4.3620%	1.0000	0.8353	0.6946	1.0000	1.000
SN22	MN	1.3974%	1.0000	0.7196	0.9065	0.9111	1.000
SN23	MS	1.7239%	1.0000	0.8461	0.9230	0.9876	1.000
SN24	MO	1.7239%	0.9697	0.8181	1.0000	0.9976	1.000
SN25	MT	2.8862%	1.0000	0.4117	0.7013	0.7058	0.1719
SN26	NE	0.8358%	1.0000	0.8437	0.6718	1.0000	1.000
SN27	NV	0.3917%	1.0000	0.6333	0.6000	0.7333	1.0000
SN28	NH	0.6007%	1.0000	0.9130	0.6956	0.9876	1.0000
SN29	NJ	4.4142%	0.5769	0.5118	1.0000	0.9977	1.0000
SN30	NM	0.6007%	1.0000	0.6739	0.8043	0.9767	1.0000
SN31	NY	7.9404%	0.8125	0.7960	1.0000	0.9876	1.0000
SN32	NC	1.3582%	1.0000	0.7788	0.7692	0.9879	1.0000
SN33	ND	0.7444%	1.0000	0.5087	0.9473	0.9871	0.6491
SN34	OH	4.3750%	1.000	0.8328	0.8746	0.9899	1.0000
SN35	OK	0.5093%	0.7949	1.0000	0.8461	0.8974	1.0000

SN36	OR	2.9776%	1.0000	0.6403	0.4210	0.9342	1.0000
SN37	PA	10.1475%	0.5534	0.9317	1.0000	0.7657	1.0000
SN38	RI	0.5876%	0.9556	1.0000	0.9333	0.9777	1.0000
SN39	SC	0.3395%	0.730	0.7692	1.0000	1.0000	1.0000
SN40	SD	0.6921%	1.0000	0.6226	0.3962	0.8113	1.0000
SN41	TN	1.0317%	0.9873	0.9240	1.0000	1.0000	1.0000
SN42	TX	1.1884%	0.3846	0.4835	1.0000	1.0000	1.0000
SN43	UT	1.319054%	1.0000	0.5049	0.3366	0.5148	0.5049
SN44	VT	1.0709%	1.0000	0.7804	0.5365	0.6219	1.0000
SN45	VA	2.6119%	0.7550	0.6600	1.0000	0.9950	1.0000
SN46	WA	4.2575%	1.0000	0.6809	0.5920	0.9190	1.0000
SN47	WV	0.5876%	1.0000	0.8666	0.6444	0.4444	1.0000
SN48	WI	6.6475%	1.0000	0.6620	0.4027	0.6149	0.6699
SN49	WY	0.8097%	1.0000	0.5967	0.8064	0.5483	0.2580

Table C.3: Description of sub-networks for 2016. $|V| = 7430$ and and scaling constant $S_c = 0.015$ for 2016

Sub-network Id	Location	Nodes (% of $\|V\|$)	Probability of Active nodes				
			Jun.	Jul.	Aug.	Sept.	Oct.
SN1	AL	0.6999 %	0.3654	0.2885	0.4423	0.5962	1.0000
SN2	AZ	0.5922 %	1.0000	0.6364	0.4545	0.3409	1.0000
SN3	AR	1.2113 %	0.2778	0.1556	0.1556	0.2111	1.0000
SN4	CA	2.7052 %	1.0000	0.6667	0.5423	0.5721	0.7363
SN5	CO	3.3513 %	1.0000	0.5783	0.4257	0.9036	1.0000

SN6	CT	1.8708 %	1.0000	0.6835	0.5108	0.6259	1.0000
SN7	DE	1.3324 %	0.2828	0.2828	0.3333	1.0000	1.0000
SN8	DC	0.6999 %	0.5000	0.4423	0.8077	1.0000	1.0000
SN9	FL	0.7402 %	0.0909	0.0909	0.0909	0.0909	1.0000
SN10	GA	1.5882 %	0.3898	0.3559	0.3305	0.4576	1.0000
SN11	ID	0.6864 %	1.0000	1.0000	0.4902	1.0000	1.0000
SN12	IL	8.1157 %	0.6683	0.5605	0.5423	1.0000	1.0000
SN13	IN	1.8708 %	0.8561	0.6547	0.5324	1.0000	1.0000
SN14	IA	0.8210 %	1.0000	0.8033	0.5738	0.9016	1.0000
SN15	KS	0.9421 %	0.7000	0.4429	0.4714	1.0000	1.0000
SN16	KY	1.1440 %	0.5647	0.5647	0.5765	1.0000	1.0000
SN17	LA	0.2153 %	0.5000	0.4375	0.5000	0.6250	1.0000
SN18	ME	0.7672 %	1.0000	1.0000	0.4561	0.6842	1.0000
SN19	MD	2.9206 %	0.7512	0.6083	0.6221	1.0000	1.0000
SN20	MA	4.8991 %	0.7555	0.6978	0.6951	1.0000	1.0000
SN21	MI	3.9435 %	1.0000	0.7782	0.8055	1.0000	1.0000
SN22	MN	1.8170 %	1.0000	0.5630	0.4444	1.0000	1.0000
SN23	MS	0.1750 %	0.4615	0.5385	0.5385	0.6923	1.0000
SN24	MO	2.1803 %	0.5741	0.4691	0.5185	1.0000	1.0000
SN25	MT	1.3190 %	1.0000	1.0000	0.8571	1.0000	0.6633
SN26	NE	0.8479 %	1.0000	0.4762	0.5079	0.8730	1.0000
SN27	NV	0.3634 %	1.0000	0.5926	0.3333	0.4074	0.8148
SN28	NH	0.5518 %	1.0000	1.0000	0.5854	1.0000	1.0000
SN29	NJ	3.9973 %	0.6801	0.4276	0.4377	1.0000	1.0000
SN30	NM	0.4980 %	1.0000	0.8108	0.6486	0.7838	1.0000
SN31	NY	6.6891 %	0.9819	0.8692	1.0000	1.0000	1.0000
SN32	NC	1.5612 %	0.5776	0.4741	0.4828	1.0000	1.0000

SN33	ND	0.6326 %	1.0000	0.5745	0.8723	1.0000	1.0000
SN34	OH	4.4145 %	0.9787	0.8780	0.7713	1.0000	1.0000
SN35	OK	1.3728 %	0.2647	0.2549	0.3039	0.4510	1.0000
SN36	OR	2.0054 %	1.0000	1.0000	0.5101	0.8322	1.0000
SN37	PA	8.7079 %	0.6955	0.5363	0.6584	1.0000	1.0000
SN38	RI	0.5787 %	1.0000	0.8372	0.4419	0.5581	0.9070
SN39	SC	0.5653 %	0.3333	0.2857	0.2619	0.4286	1.0000
SN40	SD	0.6191 %	1.0000	0.4348	0.3913	0.6304	1.0000
SN41	TN	3.3647 %	0.2680	0.2400	0.2920	0.4520	1.0000
SN42	TX	1.9246 %	0.1958	0.2587	0.3986	0.3846	1.0000
SN43	UT	1.0902 %	1.0000	0.6173	0.3704	0.5556	0.7654
SN44	VT	0.7268 %	1.0000	1.0000	0.6667	0.8148	1.0000
SN45	VA	2.7322 %	1.0000	0.6059	0.6158	0.8325	1.0000
SN46	WA	1.9112 %	1.0000	1.0000	1.0000	1.0000	1.0000
SN47	WV	0.5653 %	1.0000	0.7381	0.4762	0.6429	0.9762
SN48	WI	7.2275 %	1.0000	0.4991	0.2048	0.2365	0.5270
SN49	WY	0.4441 %	1.0000	1.0000	0.8788	1.0000	0.4545