Supplementary Material

1 Illustrative example of Thompson Sampling

We proposed to use a variant of Thompson sampling to balance learning about the system dynamics with choosing optimal decisions according to current estimated dynamics. We use a toy example to illustrate why it is necessary to explore allocations that are suboptimal according to current estimated dynamics to ensure consistent estimation of the optimal allocation strategy. Consider a system with only two locations $\mathcal{L} = \{1, 2\}$, no state information, and no temporal dependence. Assume that the potential utility at time t under allocation (a_1, a_2) is $U^{*t}(a_1, a_2) = \mu_{a_1,a_2} + \epsilon^t$ where $\epsilon^1, \epsilon^2, \ldots \sim_{i.i.d.} \text{Normal}(0, \sigma^2)$ and σ^2 is fixed and known. Assume that the decision maker can treat exactly one of the two locations at each time point so that the set of feasible allocations is $\mathcal{F} = \{(1, 0), (0, 1)\}$. Further assume that $\mu_{1,0} = 1$, and $\mu_{0,1} = 1 - \eta$ where $\eta \in (0, 1)$. Thus, the optimal allocation strategy is to select $(0, 1)$ at each time point.

At any time T, the observed data are $\mathcal{D}^T = \left\{(\boldsymbol{A}^t, U^t)\right\}_{t=1}^T$. For each feasible $\boldsymbol{a} = (a_1, a_2)$ define $n_{\boldsymbol{a}}^T = \sum_{t=1}^T 1_{\boldsymbol{A}^t = \boldsymbol{a}^t}$ and $\widehat{\mu}_{\boldsymbol{a}}^T = \left(n_{\boldsymbol{a}}^T\right)^{-1} \sum_{t=1}^T U^t 1_{\boldsymbol{A}^t = \boldsymbol{a}}$. We assume $T \geq 2$ and $n_{\boldsymbol{a}}^T \geq 1$ for all feasible $\boldsymbol{a}$. Let $(\widetilde{\mu}_{1,0}^T, \widetilde{\mu}_{0,1}^T)$ denote a draw from the posterior distribution of $(\mu_{1,0}, \mu_{0,1})$ given $\mathcal{D}^T$ then the Thompson sampling estimator is $\widehat{\pi}_{\text{TS}}^T = \arg\max_{\boldsymbol{a}} \widetilde{\mu}_{\boldsymbol{a}}^T$. This corresponds to the Thompson sampling estimator for a two-arm Gaussian bandit and thus it follows [e.g., Honda and Takemura, 2014] that $\widehat{\pi}_{TS}^T \to (0, 1)$ in probability as $T \to \infty$ (furthermore, it can be shown that under certain conditions that it attains the optimal convergence rate.)

A natural competitor is to sample from each of the feasible allocations for a fixed number of iterations and then follow the estimated optimal allocation strategy thereafter; e.g., given integer k select allocation $\boldsymbol{A}^t = (0, 1)$ for the first k time points, $\boldsymbol{A}^t = (1, 0)$ for the subsequent k time points, and then select $A^t = \arg\max_{\boldsymbol{a}} \widehat{\mu}_{\boldsymbol{a}}^t$ for $t = 2k + 1, 2k + 2, \ldots$. Note that in the context of the spatio-temporal allocation problem in the main the text, the above approach is analogous to selecting allocations according to $\arg\max_{d \in \mathcal{D}} \widehat{C}^\infty(d; \widehat{\beta}^t, \widehat{\theta}^t)$ whereas Thompson sampling selects $\arg\max_{d \in \mathcal{D}} \widehat{C}^\infty(d; \widetilde{\beta}^t, \widetilde{\theta}^t)$ (where we have set

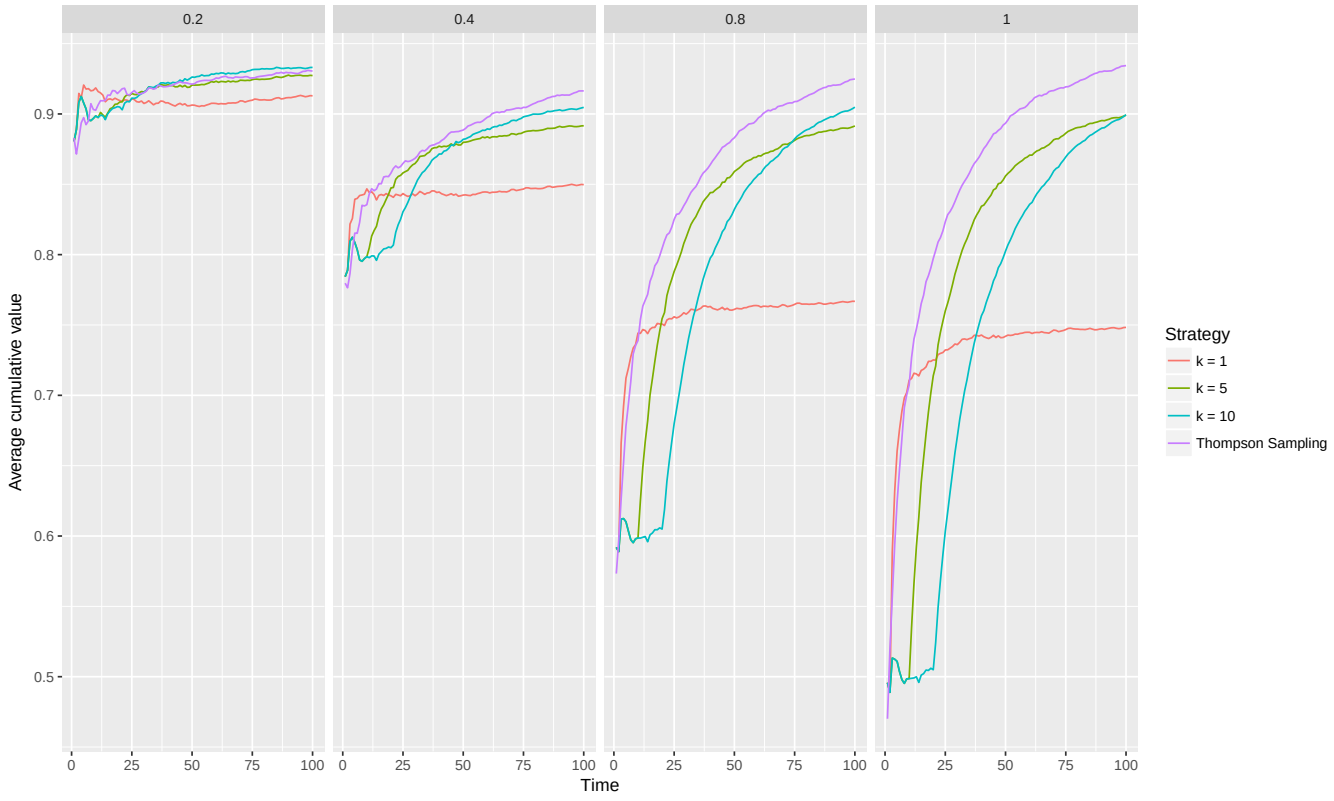

Figure 1: Average cumulative outcome for Thompson sampling and greedy allocation selection based on $k = 1, 5$, and 10 exploration steps of each allocation and a gap in means of $\eta = 0.2, 0.4, 0.8$, and 1.0 The figure illustrates that greedy action selection need not be consistent for the optimal strategy (i.e., these methods do not converge to 1). Results are based on 1000 Monte Carlo replications.

the number of extrapolation time points to ∞ as there is no extrapolation error in our toy example.) Figure 1 shows the results of a small simulation experiment comparing Thompson sampling with the greedy selection strategy across several values of k and η; reported results are based on 1000 Monte Carlo replications. As anticipated by the theory for bandit algorithms, Thompson sampling consistent (i.e., it's average cumulative reward converges to the optimal value) whereas the greedy selection methods are not.

2 Additional details for experiments using networks

To specify the generative model for simulation experiments using networks structures, we fit the network spread system dynamics model to the observed WNS data to obtain a starting point for $\boldsymbol{\theta}$. Some data points in the observed data are not possible under a network spread model, i.e., a location becomes infected while not being adjacent to an infected location. In this case, these observations are ignored to avoid a likelihood

Parameter	Posterior Mean	95% CI	Value used in simulation
θ_0	-2.26	[-2.95, -1.62]	-2.26
$\theta_{1,0}$	0.38	[0.15, 0.61]	0.38
$\theta_{1,1}$	-0.07	[-1.29, 1.15]	-0.07
$\theta_{1,2}$	0.45	[-0.69, 1.51]	0.45
$\theta_{1,3}$	1.14	[-0.27, 2.60]	1.14
$\theta_{2,0}$	-0.20	[-0.45, 0.01]	-0.20
$\theta_{2,1}$	1.06	[-0.32, 2.47]	1.06
$\theta_{2,2}$	-0.17	[-1.65, 1.22]	-0.17
$\theta_{2,3}$	-0.25	[-1.76, 1.26]	-0.25
θ_3	NA	[NA, NA]	1.72
θ_4	NA	[NA, NA]	1.72

Table 1: Estimated coefficients and 95% credible intervals (CIs) using for the network spread dynamics model fit using WNS data from 2006-2014. Rows for θ_3 and θ_4 correspond to intervention effects which are not identifiable in the WNS data as this data does not contain any interventions.

of zero. To define a network structure for observed WNS data, we define two locations as being adjacent if the boarders of the counties touch. We use $\mathrm{N}(0, 10^2)$ priors for θ_0 and independent $\mathrm{N}(0, 10)$ priors for the elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. We sample from the posterior using Metropolis sampling with Gaussian candidate distributions tuned to give acceptance probability around 0.4; 100,000 samples are generated and the first 20,000 are discarded as burn-in. Table 1 shows the estimated posterior mean, 95% credible interval, and value used in simulations for each parameter. We present plots of the estimated posterior distributions for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in Figure 2.

3 Tuning procedure for simultaneous perturbation

Let $\Delta = \{(\tau_j, \rho_j)\}_{j=1}^{J}$ denote a set of potential tuning parameters for the simultaneous perturbation algorithm. Recall that $Q^t(\beta, \theta)$ denotes the posterior for (β, θ) given $\boldsymbol{H}^t$. The tuning algorithm is displayed in Algorithm 1. In our simulation settings we set $B = 100$ and $T = 15$.

4 Generating spatial locations and network structures

In the simulation experiments, there are 3 spatial layouts and 3 network structures. For the lattice and random k-nearest neighbor setups, we describe the procedure for generating both the spatial layout and network structure. Simply ignore the adjacency matrix for a spatial layout or ignore the location coordinates for a network structure. For each $\ell = 1, \ldots, L$, let $C_\ell \in \mathbb{R}^2$ be the Cartesian coordinates of location ℓ. The

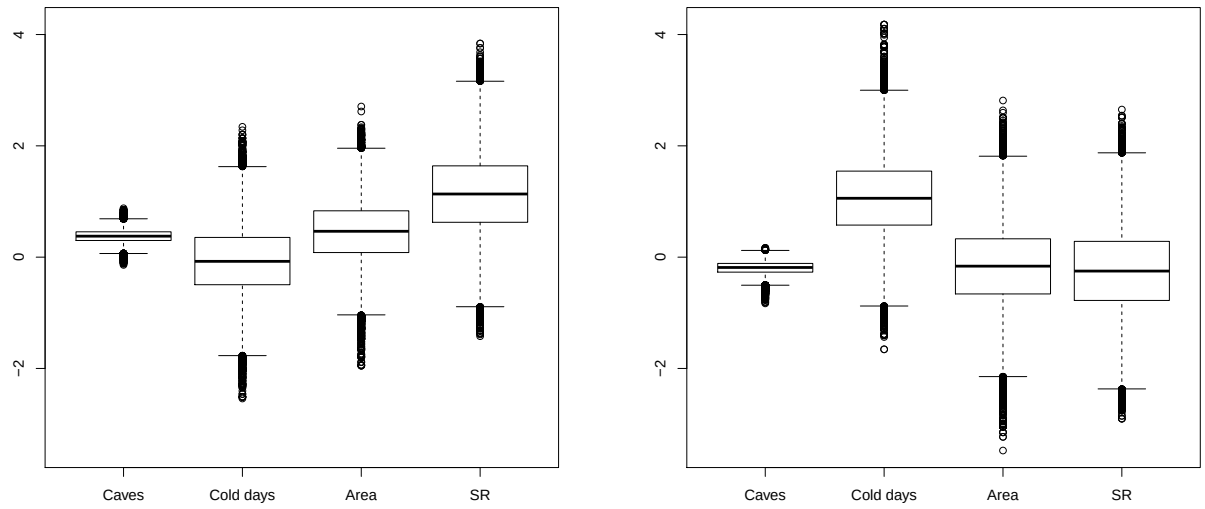

Figure 2: Posterior distribution of the regression parameters associated with uninfected ($\boldsymbol{\theta}_1$) and infected ($\boldsymbol{\theta}_2$) counties in the network spread model applied to the white-nose syndrome data. The covariates are the number of caves in the county ("caves"), the average number of days per year below 10°C ("cold days"), area in km^2 ("Area"), and species richness ("SR").

Algorithm 1: Tuning simultaneous perturbation

Input: $T < \infty$, $\boldsymbol{S}^1$, $B > 0$, $f(\boldsymbol{y}^t|\boldsymbol{s}^t, \boldsymbol{a}^t; \beta)$, $g(\boldsymbol{s}^t|\boldsymbol{s}^{t-1}, \boldsymbol{a}^{t-1}; \theta)$, $Q^t(\beta, \theta)$

for $\delta \in \Delta$ **do**
  Set $\widehat{q}_\delta = 0.0$
  for $b = 1, \ldots, B$ **do**
    Draw $(\beta^{(b)}, \theta^{(b)}) \sim Q^t(\beta, \theta)$
    Set $\boldsymbol{S}^{1(b)} = \boldsymbol{S}^1$, $\boldsymbol{H}^{1(b)} = \boldsymbol{S}^{1(b)}$
    for $r = 1, \ldots, T$ **do**
      Set $Q^{r(b)}(\beta, \theta)$ to the posterior of (β, θ) given $\boldsymbol{H}^{r(b)}$
      Compute $\pi^{r(b)} = \arg\max_{\pi \in \Pi} V^{r,T(b)}(\pi)$ using Algorithm 2 in the main paper with tuning parameters δ, and $V^{r,T(b)}(\pi)$ is computed by integrating against $Q^{r(b)}(\beta, \theta)$
      Draw $\boldsymbol{A}^{r(b)} \sim \pi^{r(b)}(\boldsymbol{S}^{r(b)})$
      Draw $\boldsymbol{Y}^{r(b)} \sim f\left\{\boldsymbol{y}^r \middle| \boldsymbol{S}^{r(b)}, \boldsymbol{A}^{r(b)}; \beta^{(b)}\right\}$
      Draw $\boldsymbol{S}^{r+1\,(b)} \sim g\left\{\boldsymbol{s}^{r+1} \middle| \boldsymbol{S}^{r(b)}, \boldsymbol{A}^{r(b)}; \theta^{(b)}\right\}$
    end
    $\widehat{q}_\delta = \widehat{q}_\delta + \sum_{r=1}^{T} \gamma^{r-1} 1^\intercal \boldsymbol{Y}^{r(b)}$
  end
end

Output: $\widehat{\delta} = \arg\max_{\delta \in \Delta} \widehat{q}_\delta$

network structures are defined by the adjacency matrix $\Omega \in \{0,1\}^{L\times L}$ where $\Omega_{\ell,\ell'} = 1$ for all $\ell = \ell'$ and the off-diagonal elements are defined in the following sections.

4.1 Lattice layout and network

The lattice network is displayed in Figure 3. First define L_1 to be the number of columns and L_2 to be the number of rows. These values are defined by

$$\{L_1, L_2\} = \underset{\substack{L_1', L_2' > 0 \\ L_1' L_2' = L \\ L_1' \leq L_2'}}{\arg\min} \quad |L_1' - L_2'|.$$

Location ℓ is positioned at $C_\ell = \{((\ell-1) \bmod L_2), \lfloor (\ell-1)/L_2 \rfloor\}$ where $\lfloor \cdot \rfloor$ denotes the floor operator. The adjacency matrix is defined as $\Omega_{\ell,\ell'} = 1$ if $\|C_\ell - C_{\ell'}\|_1 = 1$.

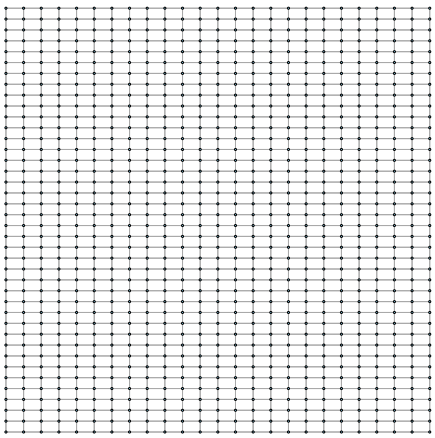

Figure 3: The lattice network with 1000 locations

4.2 Random k-nearest neighbor layout and network

The random k-nearest neighbor network is displayed in Figure 4. For each location $\ell = 1, \ldots, L$, $C_\ell \sim \{U(0,1)\}^2$. Define $E_{\ell,\ell'} = \|C_\ell - C_{\ell'}\|_2$. The set of k nearby locations to location ℓ and itself, is $N_\ell^{(k)} = \{\ell' : E_{\ell,\ell'} \leq E_{\ell,[k+1]}\}$ where $[\cdot]$ denotes the usual order statistics. Define a temporary adjacency matrix as

$\widetilde{\Omega}_{\ell,\ell'} = 1$ if $\ell \in N_{\ell'}^{(k)}$ or $\ell' \in N_{\ell}^{(k)}$ and $\widetilde{\Omega}_{\ell,\ell'} = 0$ otherwise. For the networks in the simulation experiment $k = 3$. Constructing $\widetilde{\Omega}$ in this way, may result in disjoint sub-networks. Connect these sub-networks and obtain Ω as the output of algorithm 2.

Algorithm 2: Connecting sub-networks

Input: $\widetilde{\Omega}$

Set $B = \{B_1, B_2, \ldots\}$ to be all sub-networks of $\widetilde{\Omega}$;

while $|B| > 1$ **do**

  Compute $\{\ell, \ell'\} = \underset{k \in B_{i_k}, k' \in B_{i_{k'}} : i_k \neq i_{k'}}{\arg\min} \|C_k - C_{k'}\|_2$;

  Set $\widetilde{\Omega}_{\ell,\ell'} = \widetilde{\Omega}_{\ell',\ell} = 1$;

  Set $B = \{B_1, B_2, \ldots\}$ to be all sub-networks of $\widetilde{\Omega}$;

end

Output: $\widetilde{\Omega}$

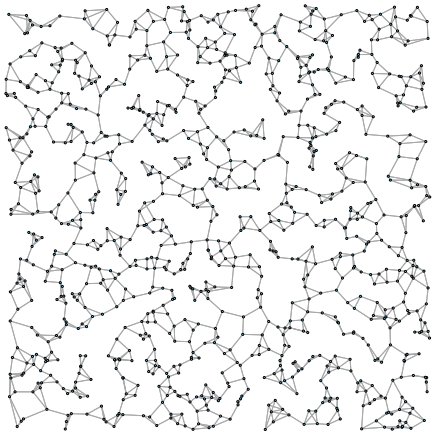

Figure 4: A random k-nearest neighbor network with 1000 locations

4.3 Scalefree ("small world") network

The scalefree network is displayed in Figure 5. In a scalefree network the proportion of nodes with k edges is proportional to $k^{-\gamma}$ for some γ. A Barabasi Network [Barabasi and Albert, 1999] ($\gamma = 3$) is used in the simulation experiment. To generate the network of size L, we will define the process as an iterative

algorithm. Define $\Omega^{(i)}$ to be the adjacency network on iteration $i = 2, \ldots, L$. A network of size 2 is the trivial network. Given a network of size $N-1$, place the additional N^{th} location adjacent to one existing location $\ell \in \{1, \ldots, N-1\}$ with probability

$$P(\Omega_{N,\ell}^{(N)} = \Omega_{\ell,N}^{(N)} = 1 | \Omega^{(N-1)}) = \frac{\sum_{j \neq \ell} \Omega_{j,\ell}^{(N-1)}}{\sum_{j,k \,:\, j \neq k} \Omega_{j,k}^{(N-1)}}.$$

Given the adjacency matrix, we use the Fruchterman-Reingold [Fruchterman and Reingold, 1991] algorithm to position the nodes in a two dimensional plane. The `R`-package `iGraph` [Csardi and Nepusz, 2006] was used to generate the scalefree network.

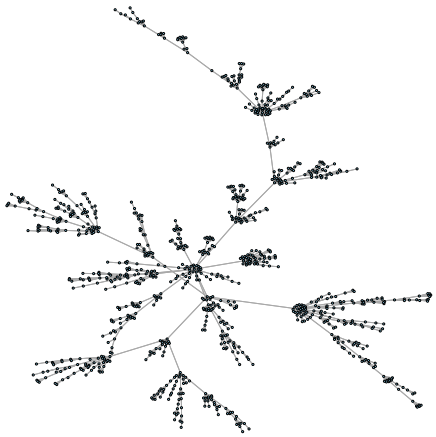

Figure 5: The scalefree network with 1000 locations

4.4 Clustered Network

The clustered network is displayed in Figure 6. A clustered network of size n is generated via the Chinese restaurant process with expected number of groups $2\log(n)$. For each new node $\ell = 1, \ldots, n$, add to a new cluster with probability $\theta_n/(\ell + \theta_n - 1)$ or add to an existing cluster of size a with probability $a/(\ell+\theta_n-1)$. Using this algorithm, the expected number of clusters in a network of size n and parameter θ is $\theta(\psi(\theta+n)-\psi(\theta))$ where $\psi(\cdot)$ is the digamma function. Thus, set θ_n to the root of $\theta(\psi(\theta+n)-\psi(\theta))-2log(n)$. If a node is starting a new cluster, then $C_\ell \sim U(0,1)^2$. If a node is joining an existing cluster b , then $C_\ell \sim N_{[0,1]}(\overline{C}_b, 0.25\boldsymbol{I})$ where $N_{[0,1]}$ is a Normal distribution truncated to $[0,1]$ and $\overline{C}_b$ is the average coordinates

for all existing nodes in cluster b.

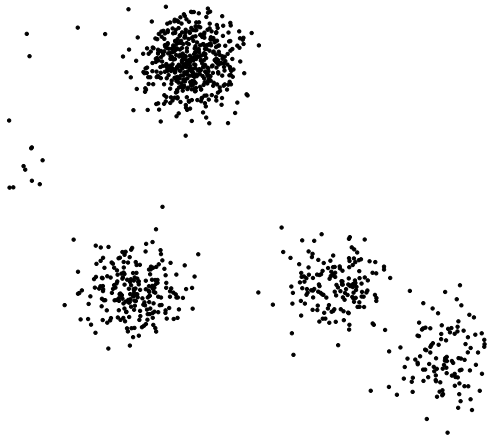

Figure 6: A clustered network with 1000 location

4.5 Covariates

The covariates are $\{X_1^\intercal, \ldots, X_L^\intercal\}^\intercal \sim N(\mathbf{0}, \Sigma)$ with $\Sigma_{(\ell-1)p+r,(\ell'-1)p+s} = \rho e^{(-\tau\|\widetilde{C}_\ell - \widetilde{C}_{\ell'}\| - \eta|r-s|)}$ for $\ell, \ell' = 1, \ldots, L$ and $r, s = 1, \ldots, p$ where $\widetilde{C}$ are centered and scaled location coordinates. The number of caves, needed to compute the gravity model, for location ℓ is defined as $Z_\ell = \lfloor X_{\ell,1} - X_{[1],1} + 1 \rfloor$. We set the parameters in Σ as $\rho = 10$, $\tau = \log(10)$, $\eta = \log(4)$ and $p = 4$. The parameters, τ and η, governing the decay in correlation are determined by

$$\begin{array}{ll} \frac{e^{(-\tau*1-\eta*0)}}{e^{(-\tau*0-\eta*0)}} & = 0.1 \\ \frac{e^{(-\tau*0-\eta*1)}}{e^{(-\tau*0-\eta*0)}} & = 0.25 \end{array}.$$

5 Tuning the generative model

To set the generative model for the simulations, we use the estimated posterior mean of θ from the observed WNS data as the starting point to provide realistic effect sizes. Let $\overline{\theta}$ be the posterior mean of θ. The parameters for the generative model are $\theta^{(g)} = \{\overline{\theta}_0 + \omega, \overline{\theta}_1, \overline{\theta}_2, \theta_3^{(g)}, \theta_4^{(g)}, \overline{\theta}_5, \overline{\theta}_6\}$. A constant, ω is added to $\overline{\theta}_0$ to force a given rate of spread. Set ω such that $\mathbb{E}_{\check{\theta}}[\mathbf{1}^\intercal \boldsymbol{Y}^T] = 0.7L$. The final component of the generative

model is defining the treatment effect sizes $\theta_3^{(g)}, \theta_4^{(g)}$. These are set such that when treating all locations $\mathbb{E}_{\theta^{(g)}}[\mathbf{1}^\intercal \boldsymbol{Y}^T] = \mathbb{E}_{\tilde{\theta}^{(g)}}[\mathbf{1}^\intercal \boldsymbol{Y}^8 + 0.05\mathbf{1}^\intercal(\boldsymbol{Y}^T - \boldsymbol{Y}^8)]$ where $\tilde{\theta}^{(g)} = \{\overline{\theta}_0 + \omega, \overline{\theta}_1, \overline{\theta}_2, 0, 0, \overline{\theta}_5, \overline{\theta}_6\}$

6 Approximating the posterior distribution of θ

Estimating the posterior distribution of θ using MCMC is computationally expensive due to the complex likelihood. To make the simulations more efficient, we approximate the posterior with the asymptotic distribution of the maximum likelihood estimator $\widehat{\theta}_{MLE}$. Under regularity conditions

$$\sqrt{n}(\widehat{\theta}_{MLE} - \theta) \longrightarrow N(0, I(\theta)^{-1}).$$

Thus the asymptotic distribution of $\widehat{\theta}_{MLE}$ is $N(\theta, I(\theta)^{-1}/n)$. The likelihood at time T is

$$L(\theta; \boldsymbol{H}^T) = \prod_{t=1}^{T-1} \prod_{\ell \in \widetilde{\mathcal{I}}^t} \left[q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)^{y_\ell^{t+1}} (1 - q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta))^{1 - y_\ell^{t+1}} \right].$$

The first derivative of the log-likelihood, $\ell(\theta)$, with respect to θ_k is

$$\frac{\partial}{\partial \theta_k} \ell(\theta; \boldsymbol{H}^T) = \sum_{t=1}^{T-1} \sum_{\ell \in \widetilde{\mathcal{I}}^t} \left[\left(\frac{y_\ell^{t+1}}{q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)} - 1 \right) \sum_{j \in \mathcal{I}^t} q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta) \frac{\partial}{\partial \theta_k} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \right].$$

The second derivative of $\ell(\theta)$ with respect to θ_k and $\theta_{k'}$ is

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_k \partial \theta_{k'}} \ell(\theta; \boldsymbol{H}^T) = \sum_{t=1}^{T-1} \sum_{\ell \in \widetilde{\mathcal{I}}^t} \Bigg[& \left(\frac{y_\ell^{t+1}}{q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)} - 1 \right) \sum_{j \in \mathcal{I}^t} \Bigg(q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta) \frac{\partial^2}{\partial \theta_k \partial \theta_{k'}} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \\
& + q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)(1 - q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)) \frac{\partial}{\partial \theta_k} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \frac{\partial}{\partial \theta_{k'}} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \Bigg) \\
& - \frac{y_\ell^{t+1}}{q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)^2} (1 - q_\ell(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)) \left(\sum_{j \in \mathcal{I}^t} q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta) \frac{\partial}{\partial \theta_k} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \right) \\
& \left(\sum_{j \in \mathcal{I}^t} q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta) \frac{\partial}{\partial \theta_{k'}} logit[q_{\ell,j}(\boldsymbol{s}^t, \boldsymbol{a}^t; \theta)] \right) \Bigg]
\end{aligned}$$

which defines $I(\theta)$. In the case of WNS, we only observe a single observation of the disease spread. Thus, our approximate posterior distribution of θ is $N(\hat{\theta}_{MLE}, I(\hat{\theta}_{MLE})^{-1})$.

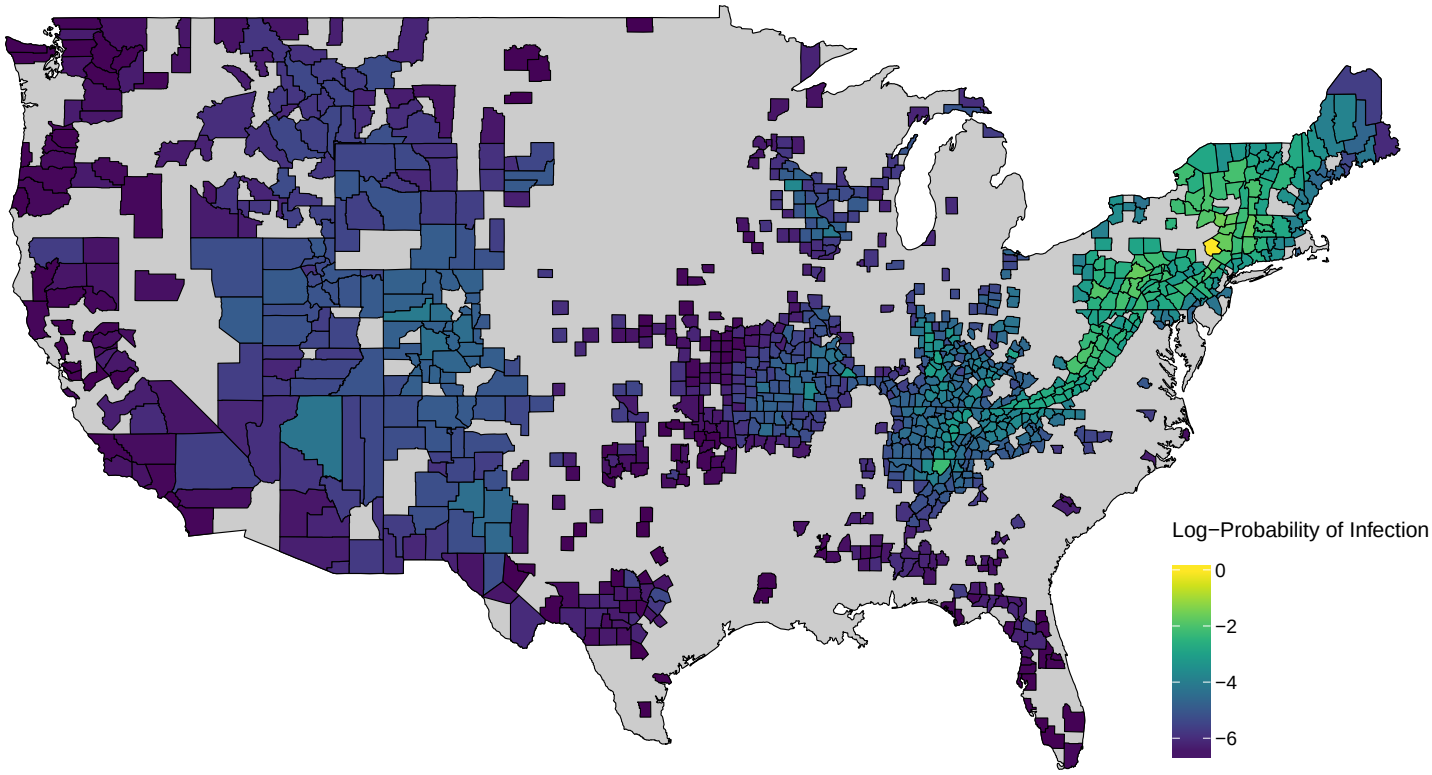

Figure 7: Estimated spread starting at 2006 and moving through 2013 for the spatial spread system dynamics model.

7 Model fit diagnostics

In this section we present figures that show the quality of fit for both the spatial and network model using the observed WNS data. First we show estimated infection spread, then we show posterior predictive checks.

7.1 Estimated infection spread

Using both estimated system dynamics models, we estimate the posterior probability of each location being infected by 2013 from the original observed data point in 2006. Figure 7 shows average infections under the spatial model and Figure 8 shows average infections under the network model.

Additionally, we show two more figures for estimated infections using out-of-sample model fitting. We estimate the system dynamics models using only data from 2006 to 2011. Then starting from 2011, simulate

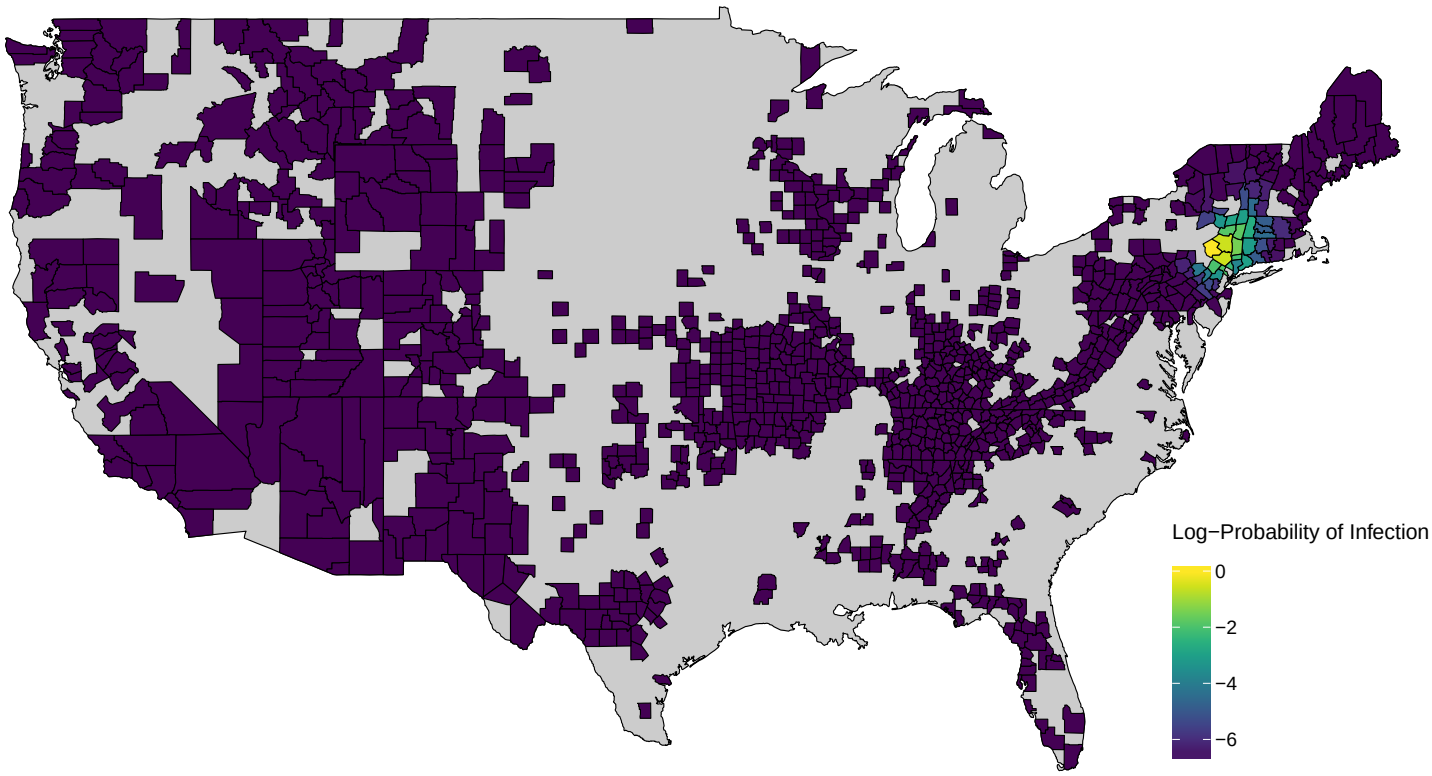

Figure 8: Estimated spread starting at 2006 and moving through 2013 for the network spread system dynamics model.

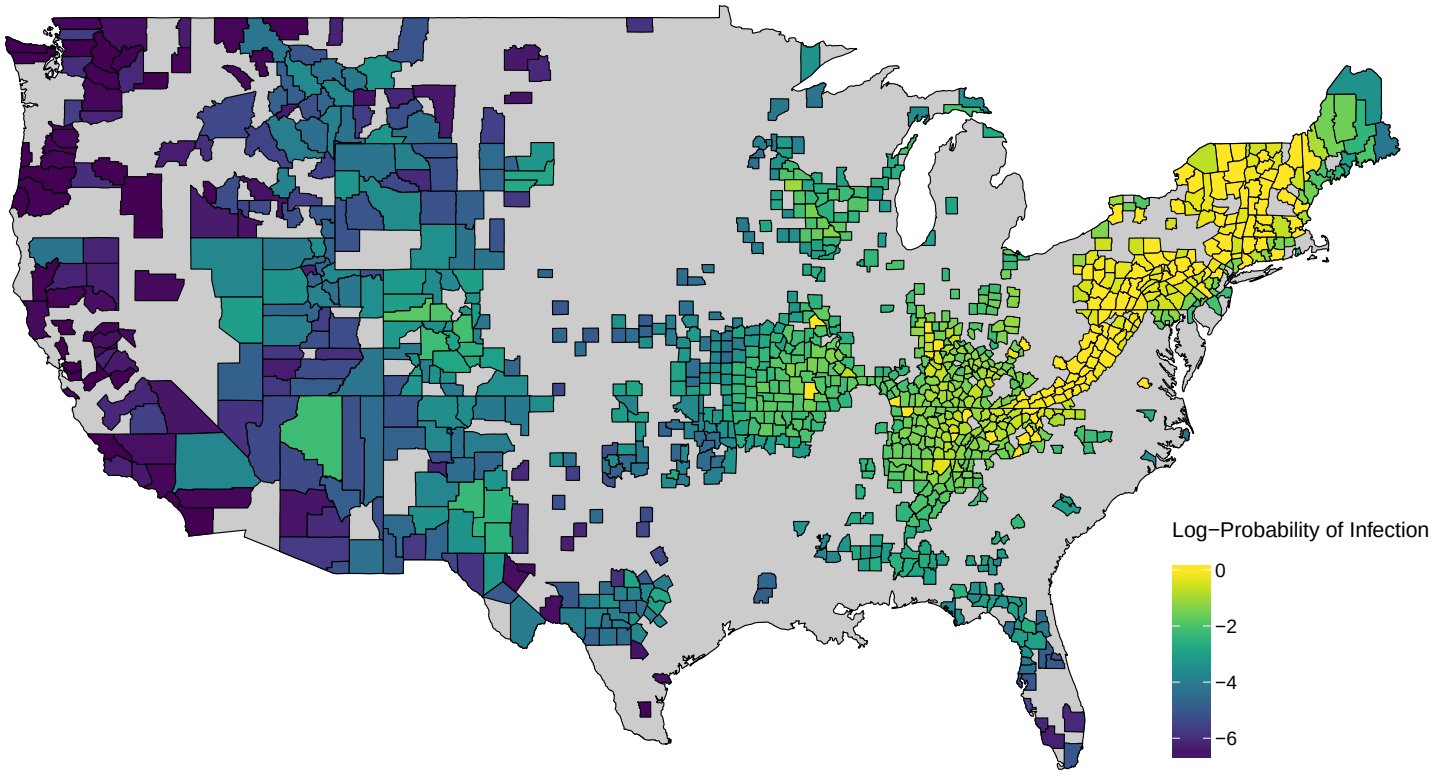

Figure 9: Estimated spread starting at 2011 and moving through 2013 for the spatial spread system dynamics model.

the final two years. Figure 9 shows average infections for the final two years under the spatial model and Figure 10 shows average infections for the final two years under the network model.

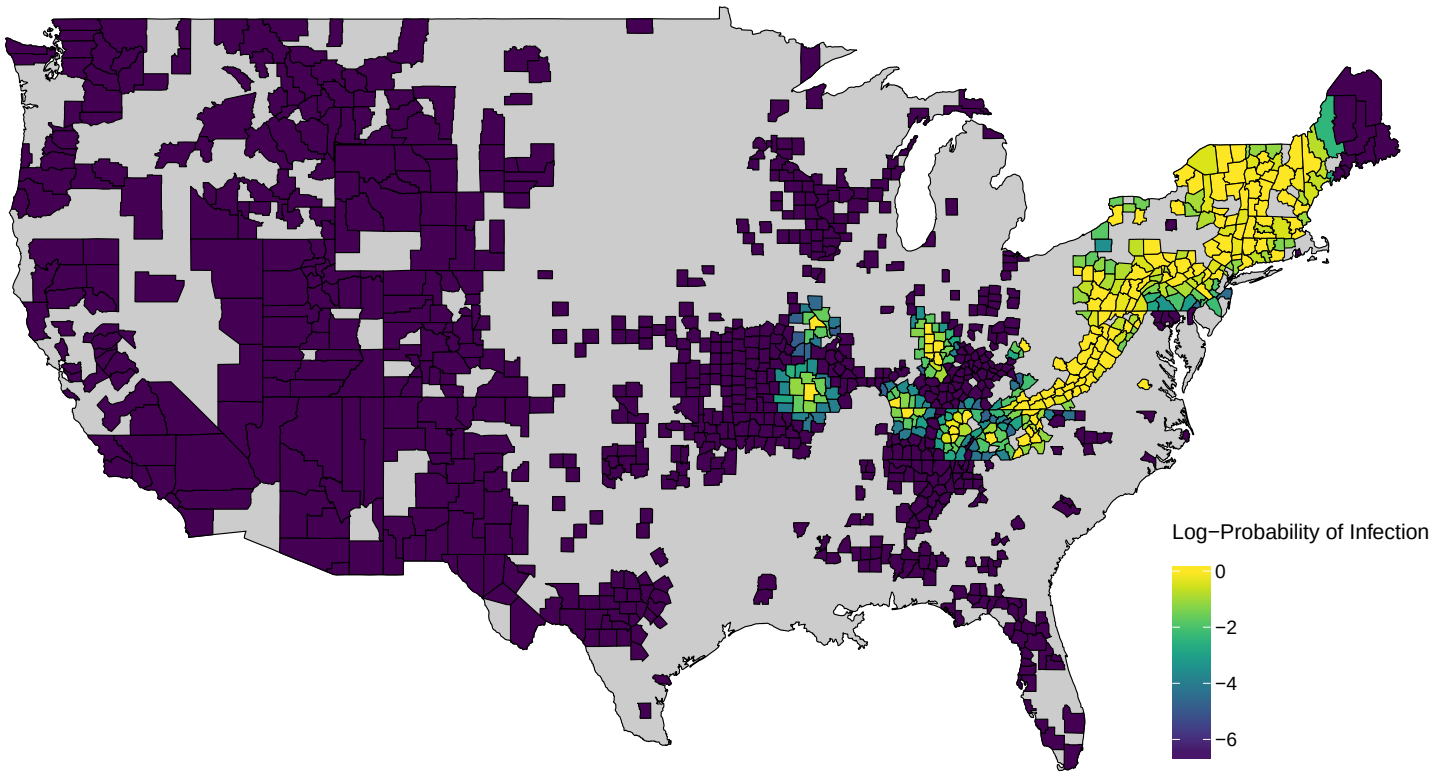

Figure 10: Estimated spread starting at 2011 and moving through 2013 for the network spread system dynamics model.

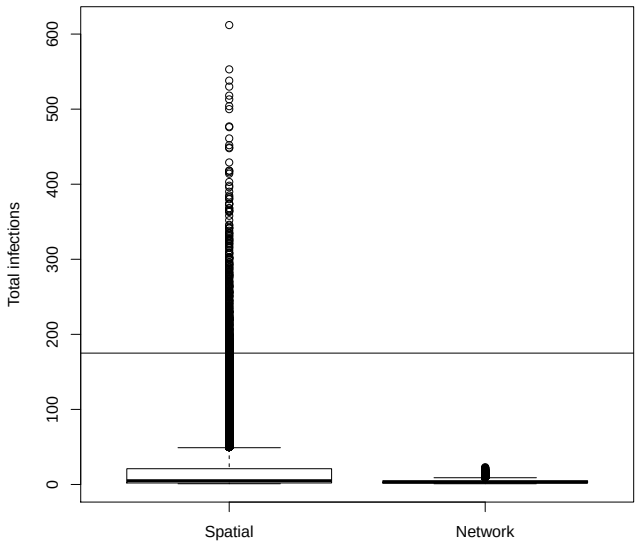

Figure 11: Posterior predictive check for the total number of infections.

7.2 Posterior predictive checks

Figures 11 to 27 show posterior predictive checks for both spatial and network models. The calculated statistic is labeled on the y-axis. Each box plot is the estimated posterior distribution of the statistic using the model fit to WNS data. The horizontal line represents the value of the observed statistic.

Figures 28 to 39 show out-of-sample posterior predictive checks for both spatial and network models. Data from years 2006 to 2011 was used to fit the models. Each statistic is calculated only using data from the final two years, 2012 and 2013. The calculated statistic is labeled on the y-axis. Each box plot is the estimated posterior distribution of the statistic using the model fit to WNS data. The horizontal line represents the value of the observed statistic.

References

Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439): 509–512, 1999. doi: 10.1126/science.286.5439.509.

Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*,

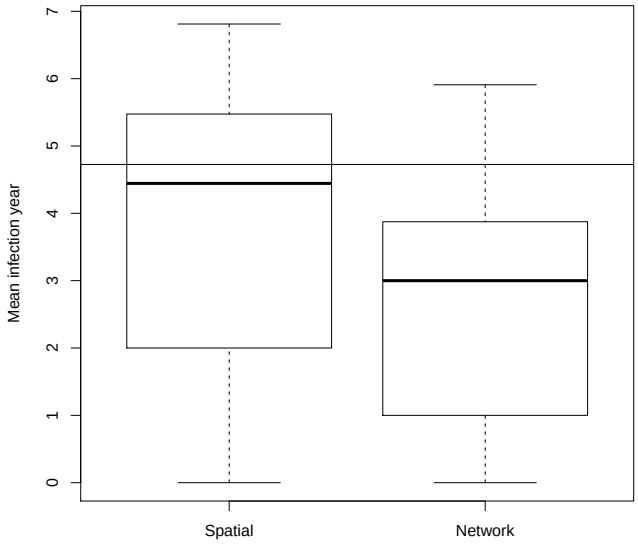

Figure 12: Posterior predictive check for mean year of infection.

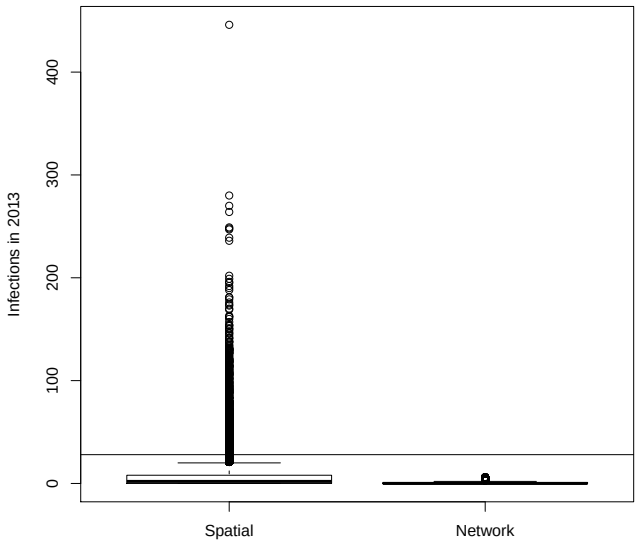

Figure 13: Posterior predictive check for the number of infections in 2013.

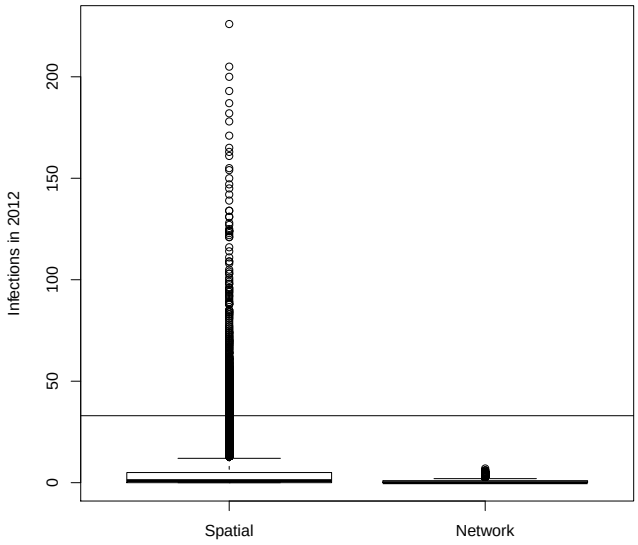

Figure 14: Posterior predictive check for the number of infections in 2012.

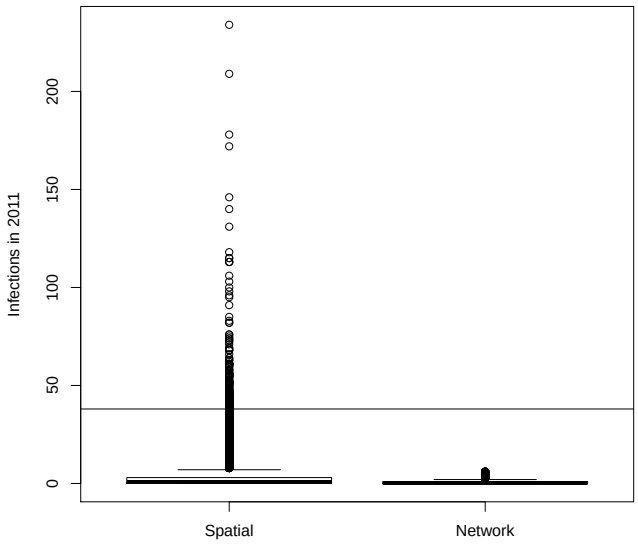

Figure 15: Posterior predictive check for the number of infections in 2011.

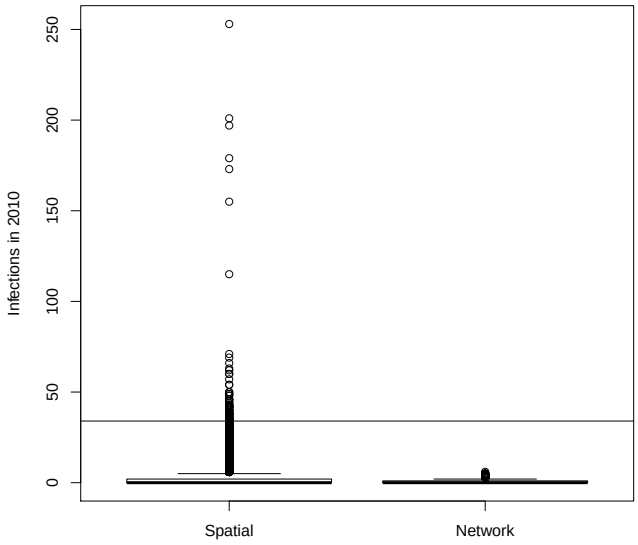

Figure 16: Posterior predictive check for the number of infections in 2010.

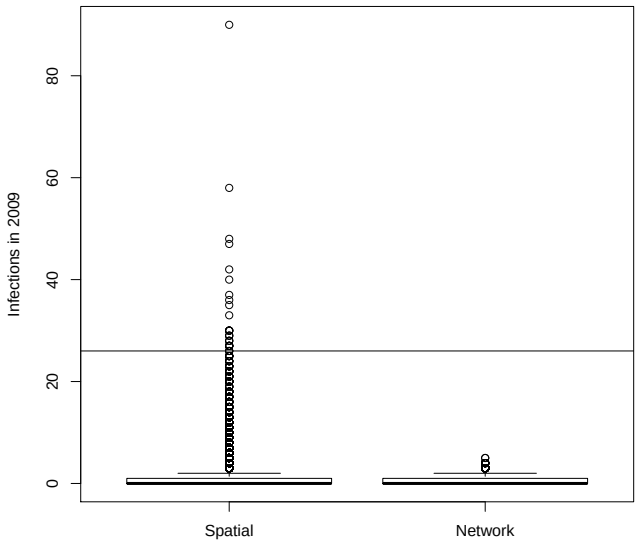

Figure 17: Posterior predictive check for the number of infections in 2009.

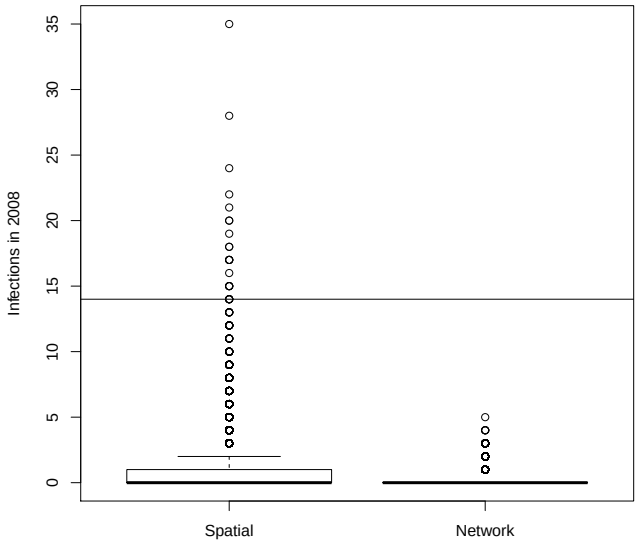

Figure 18: Posterior predictive check for the number of infections in 2008.

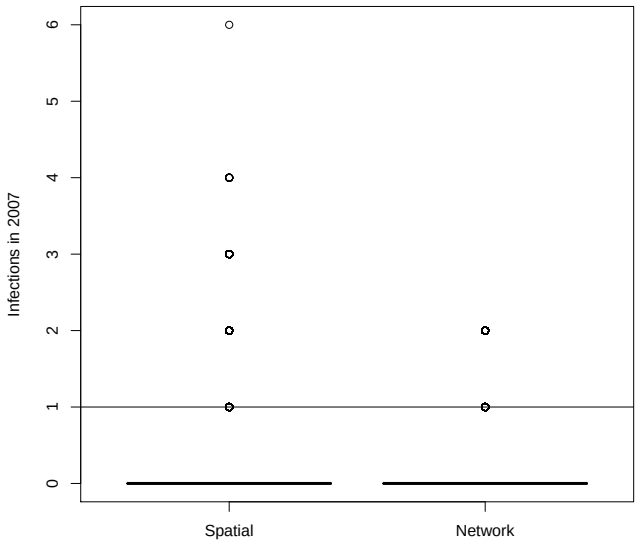

Figure 19: Posterior predictive check for the number of infections in 2007.

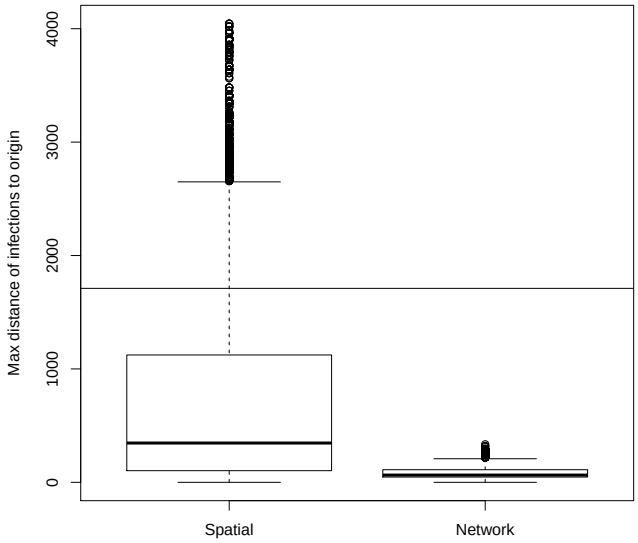

Figure 20: Posterior predictive check for maximum distance from starting location to final set of infected locations.

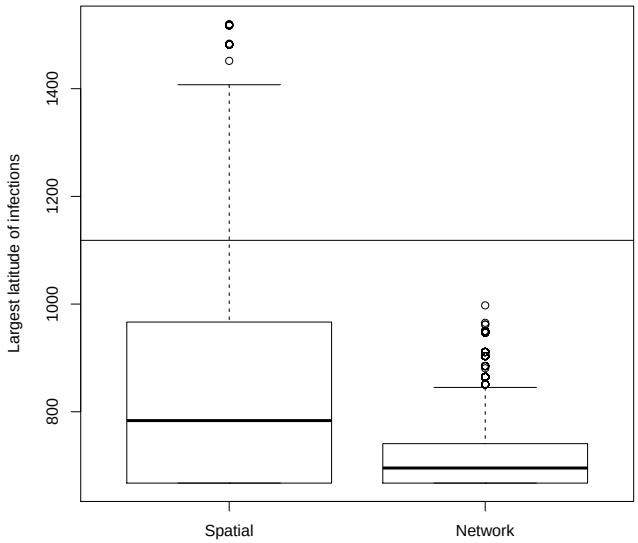

Figure 21: Posterior predictive check for maximum difference in latitude from the starting location to final set of infected locations.

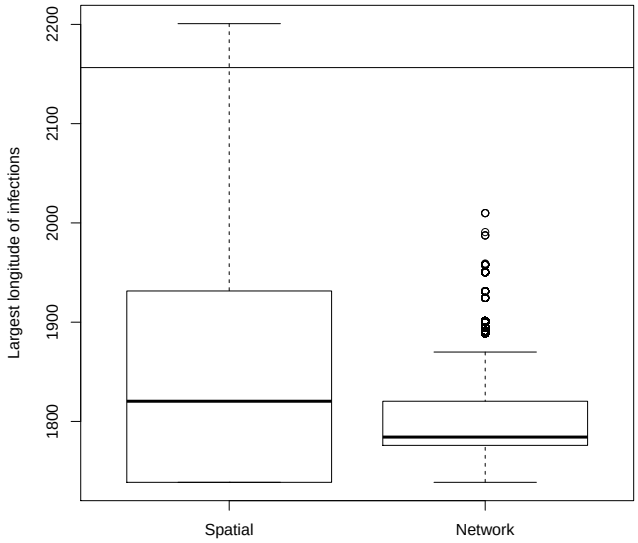

Figure 22: Posterior predictive check for maximum difference in longitude from the starting location to final set of infected locations.

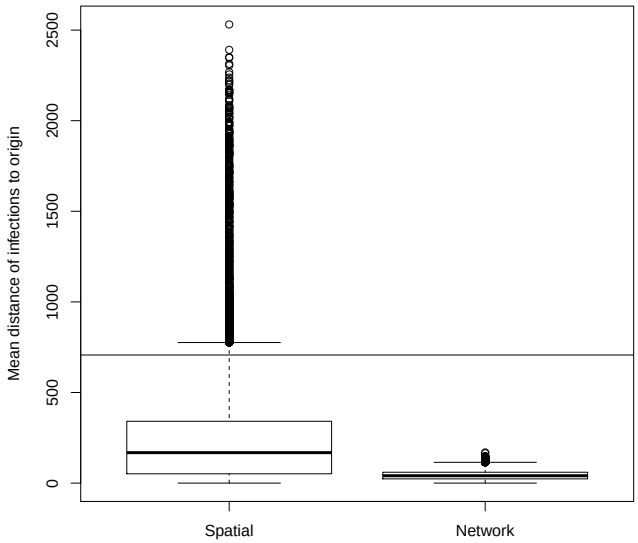

Figure 23: Posterior predictive check for mean distance from the starting location to final set of infected locations.

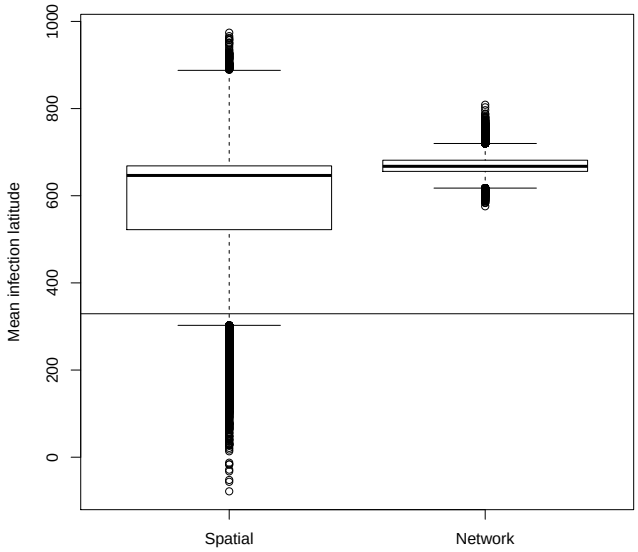

Figure 24: Posterior predictive check for mean difference in latitude from the starting location to final set of infected locations.

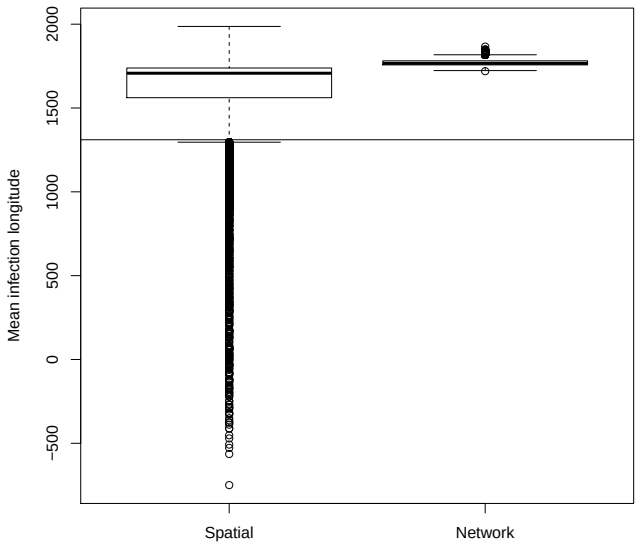

Figure 25: Posterior predictive check for mean difference in longitude from the starting location to final set of infected locations.

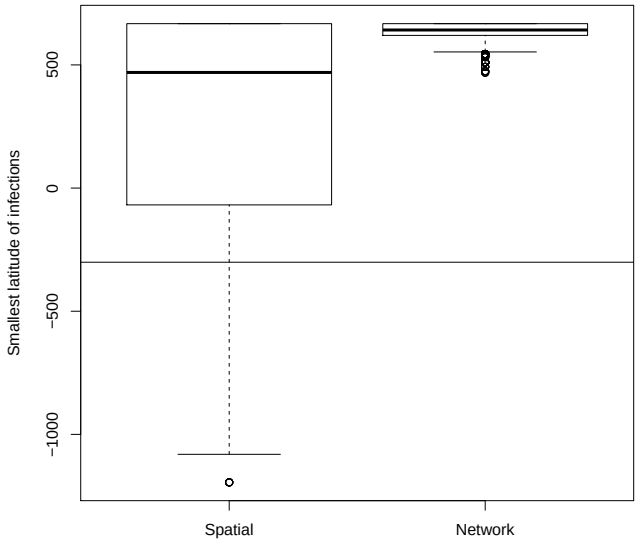

Figure 26: Posterior predictive check for minimum difference in latitude from the starting location to final set of infected locations.

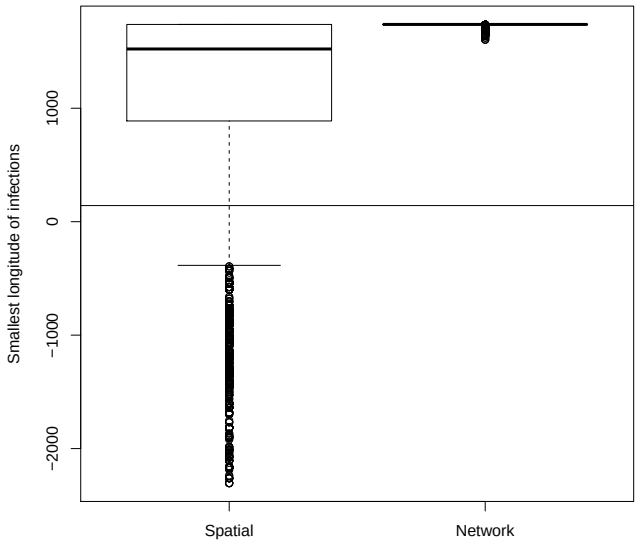

Figure 27: Posterior predictive check for minimum difference in longitude from the starting location to final set of infected locations.

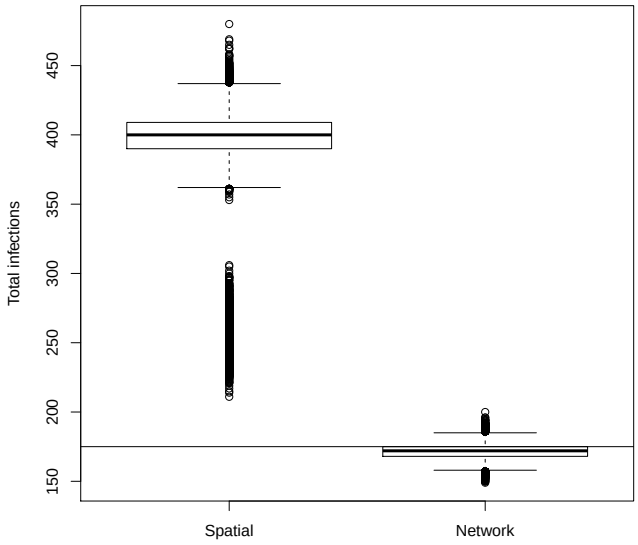

Figure 28: Posterior predictive check for the total number of infections using out of sample statistics.

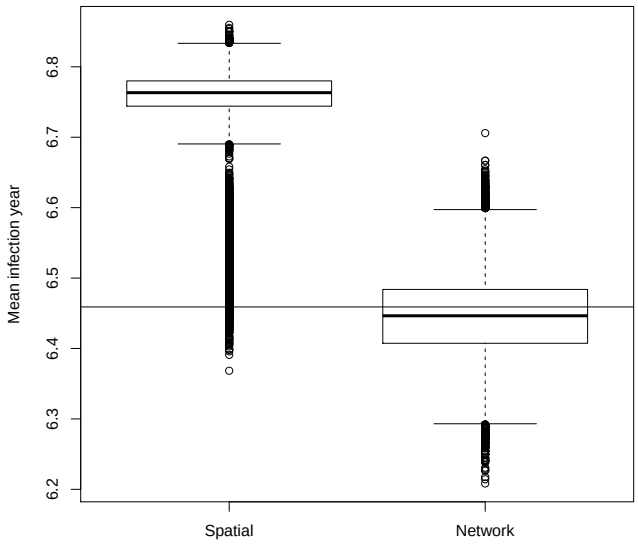

Figure 29: Posterior predictive check for the mean infection year using out of sample statistics.

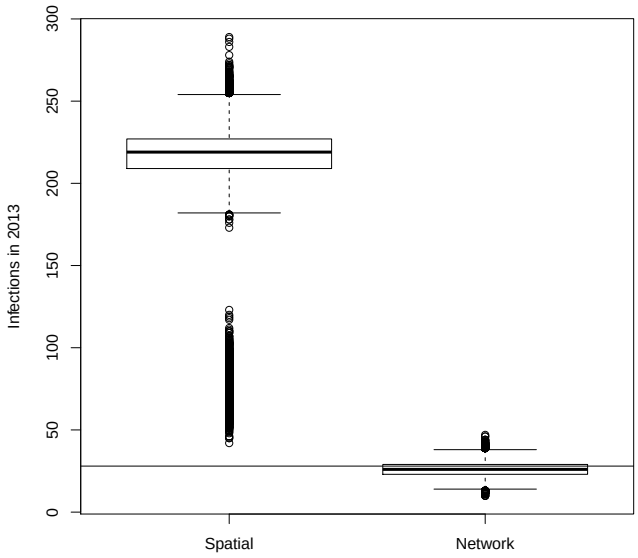

Figure 30: Posterior predictive check for the number of infections in 2013 using out of sample statistics.

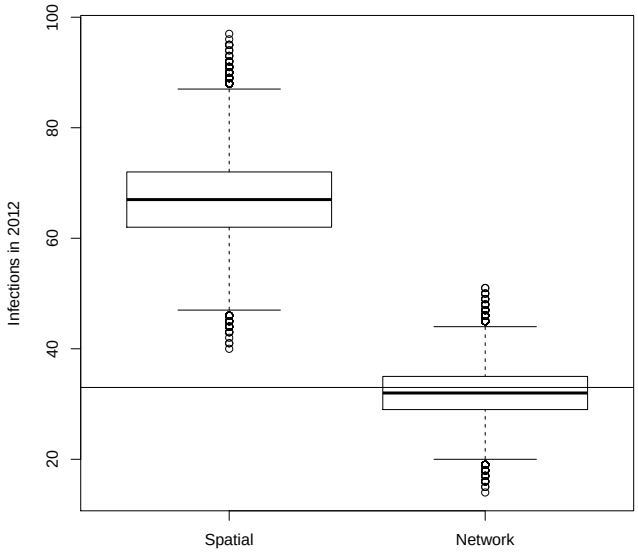

Figure 31: Posterior predictive check for the number of infections in 2012 using out of sample statistics.

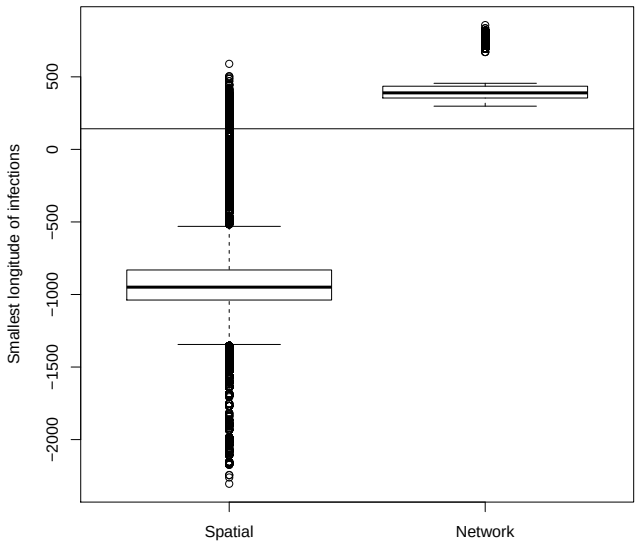

Figure 32: Posterior predictive check for the minimum difference in longitude from the starting location to final set of infected locations using out of sample statistics.

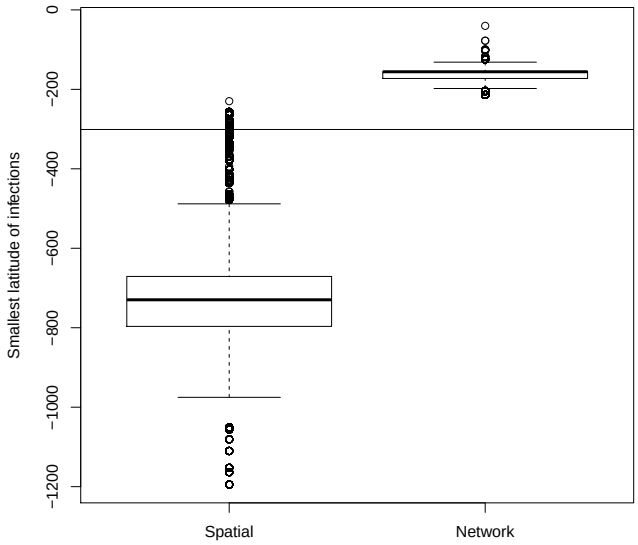

Figure 33: Posterior predictive check for the minimum difference in latitude from the starting location to final set of infected locations using out of sample statistics.

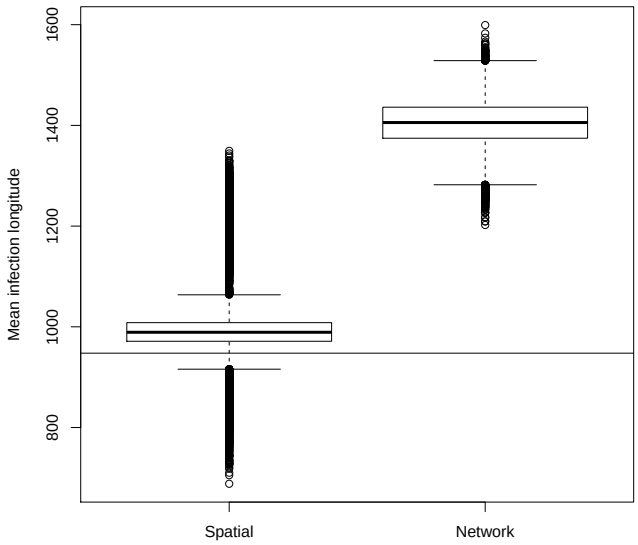

Figure 34: Posterior predictive check for the mean difference in longitude from the starting location to final set of infected locations using out of sample statistics.

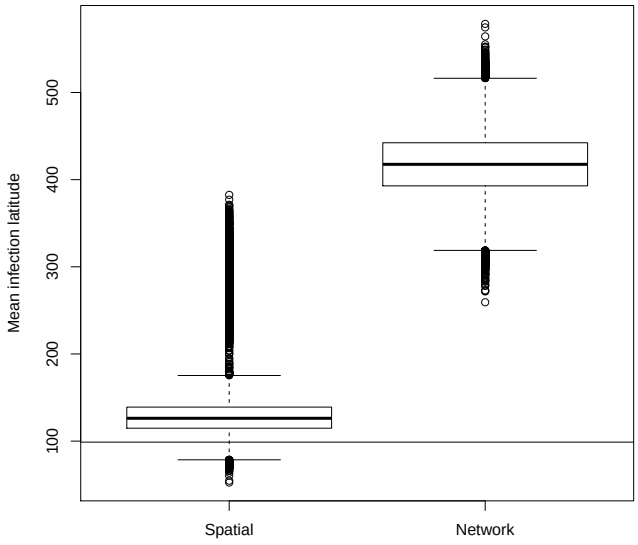

Figure 35: Posterior predictive check for the mean difference in latitude from the starting location to final set of infected locations using out of sample statistics.

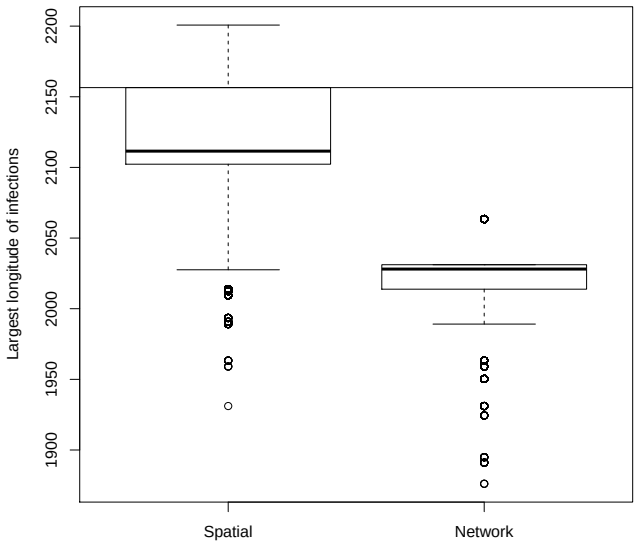

Figure 36: Posterior predictive check for the max difference in longitude from the starting location to final set of infected locations using out of sample statistics.

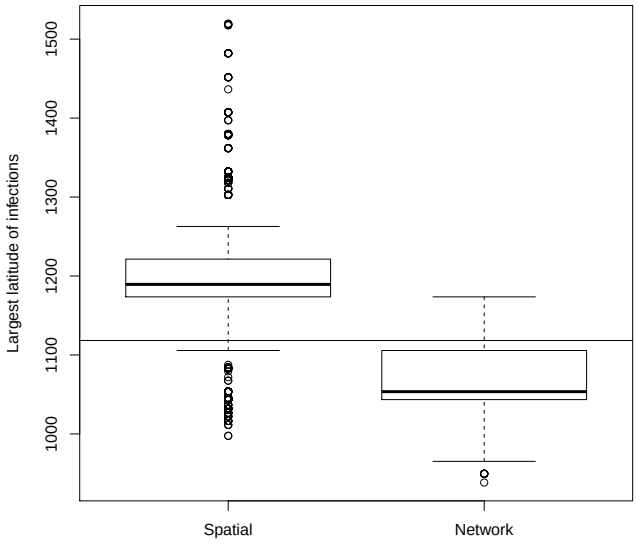

Figure 37: Posterior predictive check for the max difference in latitude from the starting location to final set of infected locations using out of sample statistics.

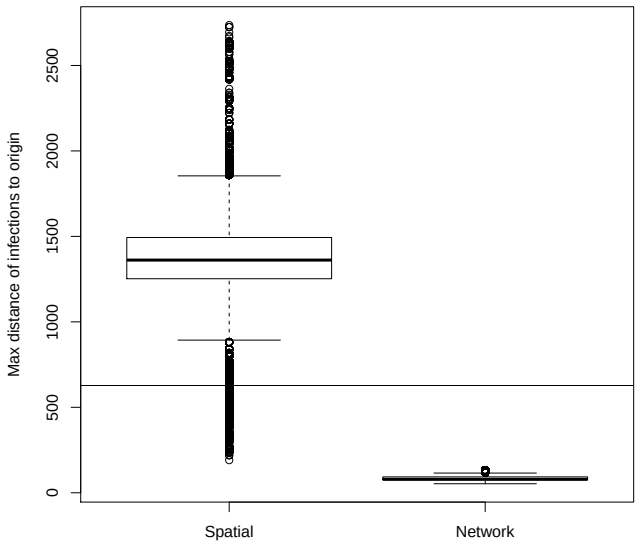

Figure 38: Posterior predictive check for the max distance from the starting location to final set of infected locations using out of sample statistics.

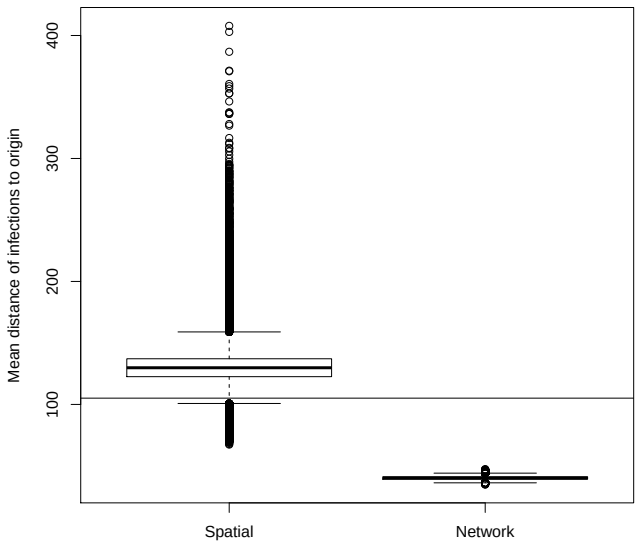

Figure 39: Posterior predictive check for the mean distance from the starting location to final set of infected locations using out of sample statistics.

Complex Systems:1695, 2006. URL http://igraph.org.

Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991. ISSN 1097-024X. doi: 10.1002/spe.4380211102. URL http://dx.doi.org/10.1002/spe.4380211102.

Junya Honda and Akimichi Takemura. Optimality of thompson sampling for gaussian bandits depends on priors. In *AISTATS*, pages 375–383, 2014.