Evidence-based controls for epidemics using spatio-temporal stochastic models in a Bayesian framework

Hola K. Adrakey, George Streftaris, Nik J. Cunniffe, Tim R. Gottwald, Christopher A. Gilligan, Gavin J. Gibson

J. R. Soc. Interface

Electronic Supplementary Material (ESM)

1 Parameter estimation

In this section we provide the description of the MCMC algorithm used in the main document to sample from the joint posterior distribution of the model parameters and the epidemic trajectories (Gibson and Renshaw, 1998; O'Neill and Roberts, 1999; Streftaris and Gibson, 2004; Neri et al., 2014; Parry et al., 2014; Lau et al., 2015). In summary, we construct a Markov chain with state vector $(\boldsymbol{\theta}, \underline{x}(T))$ whose stationary distribution is $\pi(\boldsymbol{\theta}, \underline{x}(T)|y)$.

1.1 Complete data likelihood

We first define the likelihood in the 'complete-data' setting where the observations are $\underline{x}(T)$ and comprise the precise times of infection, and identify the infected individuals, for all infections occurring in the interval $[t_0, T]$ for some $T \geq t_{obs} - \Delta$. We assume t_0 to be the time at which the primary source of infection becomes active. Let $\mathcal{I}$ denote the set of infected hosts, $\tilde{\mathcal{I}}$ its complement and, for $i \in \mathcal{I}$, let t_i denote its infection time. Then a complete-data likelihood can be constructed as:

$$\pi(\underline{x}(T)|\boldsymbol{\theta}) = \prod_{i \in \mathcal{I}} \lambda_i(t_i^-) \exp\left(-\int_{t_0}^{t_i} \lambda_i(u)du\right) \times \prod_{\boldsymbol{s} \in \tilde{\mathcal{I}}} \exp\left(-\int_{t_0}^{T} \lambda_s(u)du\right)$$

where $\lambda_i(t) = \lambda_i(t, \underline{x}(T), \theta)$ denotes the infectious challenge presented to i at time t, under the realisation $\underline{x}(T)$, for $t \in [t_0, T]$.

For computational purposes it is helpful to write this likelihood using the approach of **?** and **?**. First assign the 'notional' infection time for all $j \in \tilde{\mathcal{I}}$ to be $t_j = T$. Then we can write

$$\pi(\underline{x}(T)|\boldsymbol{\theta}) = \prod_{i \in \mathcal{I}} \lambda_i(s_i^-) \exp\left(-\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{I} \cup \tilde{\mathcal{I}}} A_{ij}(t_j - t_i)\mathbb{1}_{t_j > t_i} + \epsilon \sum_{j=1}^{N} t_j\right) \tag{1}$$

where $A_{ij} = \beta K(d_{ij}, \alpha)$ is the infectious pressure of host i on j and $\mathbb{1}$ the indicator function.

1.2 MCMC updating of θ

We update parameters using a single-component Metropolis algorithm, since the posterior conditional distribution of each one of the parameters does not have a convenient form, by proposing and accepting or rejecting changes to the current values $(\theta_1, \theta_2, \theta_3)$ of α, β and ϵ, respectively. To update each parameter we propose a new value from a normal distribution with mean the current value. More precisely, the new parameter θ_k', $k = 1, 2, 3$ is proposed as follows:

$$\theta_k' = \theta_k + u, \text{ where } u \sim N(0, \sigma_k^2). \tag{2}$$

If θ_k' lies in the range of its uniform prior, it is then accepted with probability

$$\rho = \min\left\{1, \frac{\pi(\boldsymbol{\theta}', \underline{x}(T)|\mathbf{y})}{\pi(\boldsymbol{\theta}, \underline{x}(T)|y)}\right\} = \min\left\{1, \frac{\pi(\boldsymbol{\theta}')\pi(\underline{x}(T)|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})\pi(\underline{x}(T)|\boldsymbol{\theta})}\right\} \tag{3}$$

where $\boldsymbol{\theta}'$ is the vector parameter $\boldsymbol{\theta}$ with θ_k replaced by θ_k'. Note that σ_k is a positive parameter which is tuned for each parameter separately to ensure the chain mixes well. Otherwise θ_k is left unchanged by the update.

1.3 Updating $\underline{x}(T)$

To update $\underline{x}(T)$ we need to take account of the fact that when $T > t_{obs}$, the number of infection events in $\underline{x}(T)$ is not specified by $\mathbf{y}$. This is overcome using a standard reversible-jump algorithm.

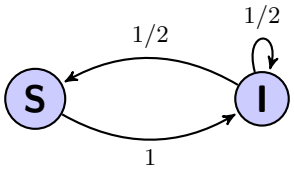


Figure 1: State diagram for the infection times to show the state transitions.

Given the current state of an individual, we propose changes to the infection times (that would leave the observation y unchanged) by adding, moving or deleting its infection time (see Figure 1). This is done as follows:

Updating $\underline{x}(T)$ using Reversible-Jump MCMC ($T > t_{obs} - \Delta$)

i) Choose an individual j in the population.

ii) If j is symptomatic at t_{obs} so that $t_j < t_{obs} - \Delta$, propose a new infection time

$$t' \sim U\left(\tau_{f(j)-1} - \Delta, \tau_{f(j)} - \Delta\right) \tag{4}$$

where $f(i)$ indexes the assessment time at which j was first symptomatic. Accept t' with probability

$$\rho = \min\left\{1, \frac{\pi(\boldsymbol{\theta}, \underline{x}'(T)|\mathbf{y})}{\pi(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})}\right\} = \min\left\{1, \frac{\pi(\underline{x}'(T)|\boldsymbol{\theta})}{\pi(\underline{x}(T)|\boldsymbol{\theta})}\right\} \tag{5}$$

where $\underline{x}'(T)$ denotes the trajectory with t_j replaced by t'.

iii) If j is infected at T (but not symptomatic at t_{obs}):

(a) with probability $p = 1/2$ move its infection time $t' \sim U[t_{obs} - \Delta, T]$ and t' with probability given by Equation 5,

(b) with probability $p = 1/2$ delete its infection time . The acceptance probability is then given by

$$\rho = \min\left\{1, \frac{2}{(T - t_{obs} + \Delta)} \frac{\pi(\underline{x}'(T)|\boldsymbol{\theta})}{\pi(\underline{x}(T)|\boldsymbol{\theta})}\right\}. \tag{6}$$

iv) If j is not infected at T, propose an infection time $t' \sim U[t_{obs} - \Delta, T]$. Accept t' with probability

$$\rho = \min\left\{1, \frac{(T - t_{obs} + \Delta)}{2} \frac{\pi(\underline{x}'(T)|\boldsymbol{\theta})}{\pi(\underline{x}(T)|\boldsymbol{\theta})}\right\}. \tag{7}$$

2 Evaluation of the function $h(\theta, \underline{Q}, d, T)$

Here, we describe the algorithm that evaluates the trajectory $\underline{x}$ introduced in Section 2.4 in the main document using Sellke thresholds (Sellke, 1983). In reality, to embed a control strategy (or testing regime in this case) into an epidemic process generated through the Sellke construction, we first generate realisations of the Sellke thresholds for each host and the model parameter from $\pi(\boldsymbol{\theta}, \underline{Q}|\mathbf{y})$. Then, starting from $t = 0$, we compute a potential infection times for each host based on its Sellke threshold and the infectious challenge it currently experiences. The next event is then chosen to be either a test or an infection, of the individual with the smallest potential infection time, depending on which event type occurs first. The disease status is then changed according to the event, infectious challenges and Sellke thresholds are updated, and the process continues until no further event can occur or until a defined stopping criterion is reached.

Calculation of $\underline{x} = h(\boldsymbol{\theta}, \underline{Q}, \boldsymbol{d}, T)$ for spatio-temporal SI model with primary and secondary infection.

Notation

- Let $\boldsymbol{d}$ denote the control which involves considering a specified subset Γ of hosts at time t_C and removing them if infected. Note that Γ is determined by the choice of prioritisation measure and events occurring up to $t_{obs} < t_C$.
- $\underline{Q} = (Q_1, ..., Q_N)$, the Sellke thresholds for the individuals.
- $\boldsymbol{\theta} = (\epsilon, \beta, \alpha)$, the primary and secondary infection rates.

Calculation

- Set $t = 0$, set of infections $I(t) = \emptyset$, infectious challenge to individual i, $\lambda_i(t) = \left(\beta \sum_{j \in I(t)} K(d_{ji}, \alpha) + \epsilon\right)$. Set the initial state of each individual, $s_i = 0, i = 1, ..., N$, and time till next transition (for current infectious challenge) $r_i = Q_i/\lambda_i(t)$. Set the final time T. Set initial waiting time till implementation of control to be t_C.
- While $t < T$ do the following {

  Identify next event: Let $t^* = \min\{r_1, ..., r_N, t_C\}$.

  If $t^* = r_j$ (event is an infection), implement event.

  - set $s_j = 1$, $I(t) = I(t) \cup \{j\}$.

  If $t^* = t_C$, remove any infected members of the control set C.

  Finally update current time, remaining Sellke thresholds, infectious challenges, and time till control in response to the event that has been implemented.

  - $t = t + t^*$,
  - $R_i' = \lambda_i(t) = \left(\beta \sum_{j \in I(t)} K(d_{ji}, \alpha) + \epsilon\right)$.
  - If $s_i = 0$, $Q_i = Q_i - R_i' t^*$, $r_i = Q_i/\lambda_i(t)$.
  - Update time remaining till implementation of control $t_C = t_C - t^*$ (if not already implemented). }

3 Example of uniformly distributed host population

3.1 Model assumptions

We consider epidemics governed by SI models in a population of $N = 1000$ uniformly distributed hosts in a $0.75 \times 0.75\text{km}^2$ square region using an exponential kernel $K(d, \alpha) = \frac{1}{2\pi d\alpha} \exp(-d/\alpha)$. We set $\alpha = 0.08$km, $\beta = 7.10^{-6}\text{days}^{-1}\text{km}^2$ and $\epsilon = 5.10^{-5}\text{days}^{-1}$ for the simulation. Snapshots are taken over a period of 360 days at 30-day

intervals as shown in Figure (2). As discussed in the main text, the data **y** specify the location of each host and a period of time where each symptomatic host is infected. By the end of the observation, there are 128 symptomatic hosts detected while 153 infections are undetected (cryptic).

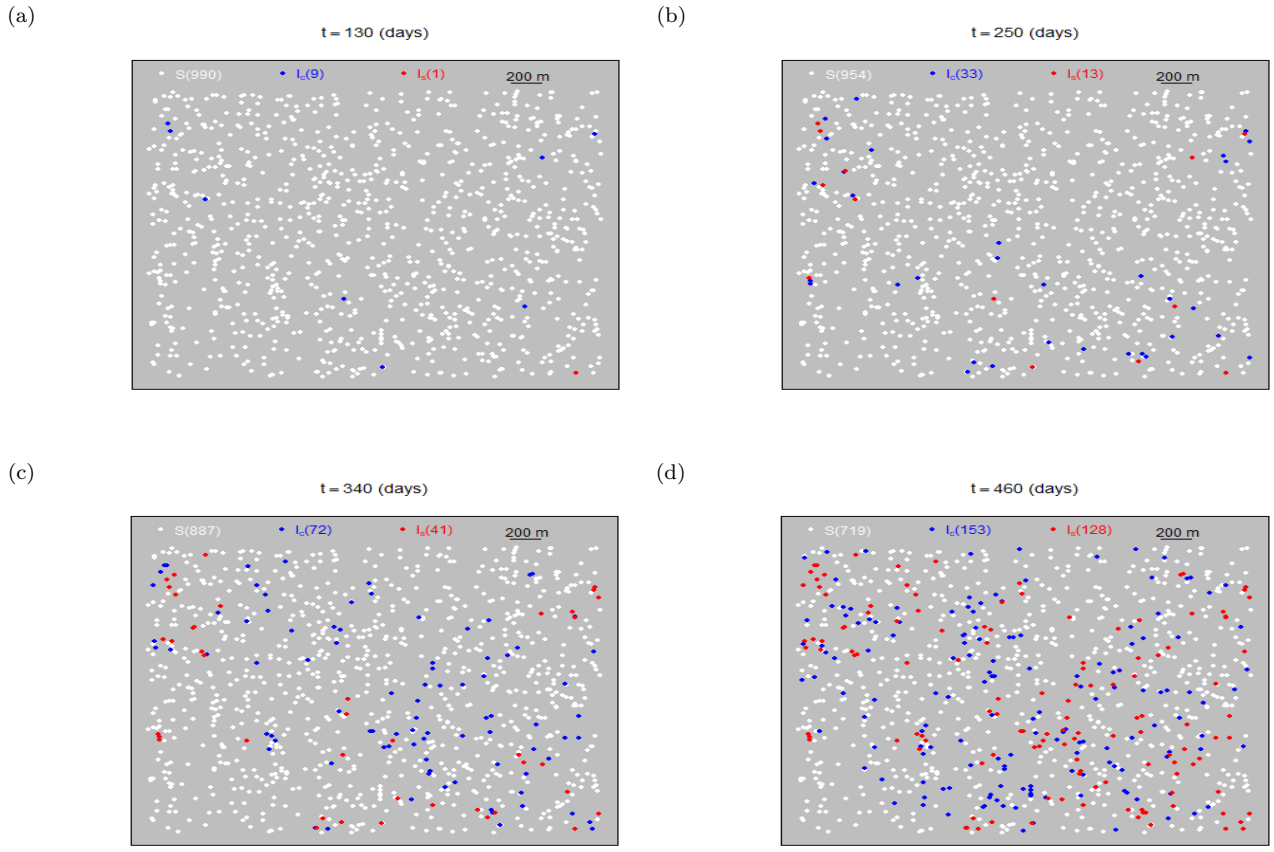


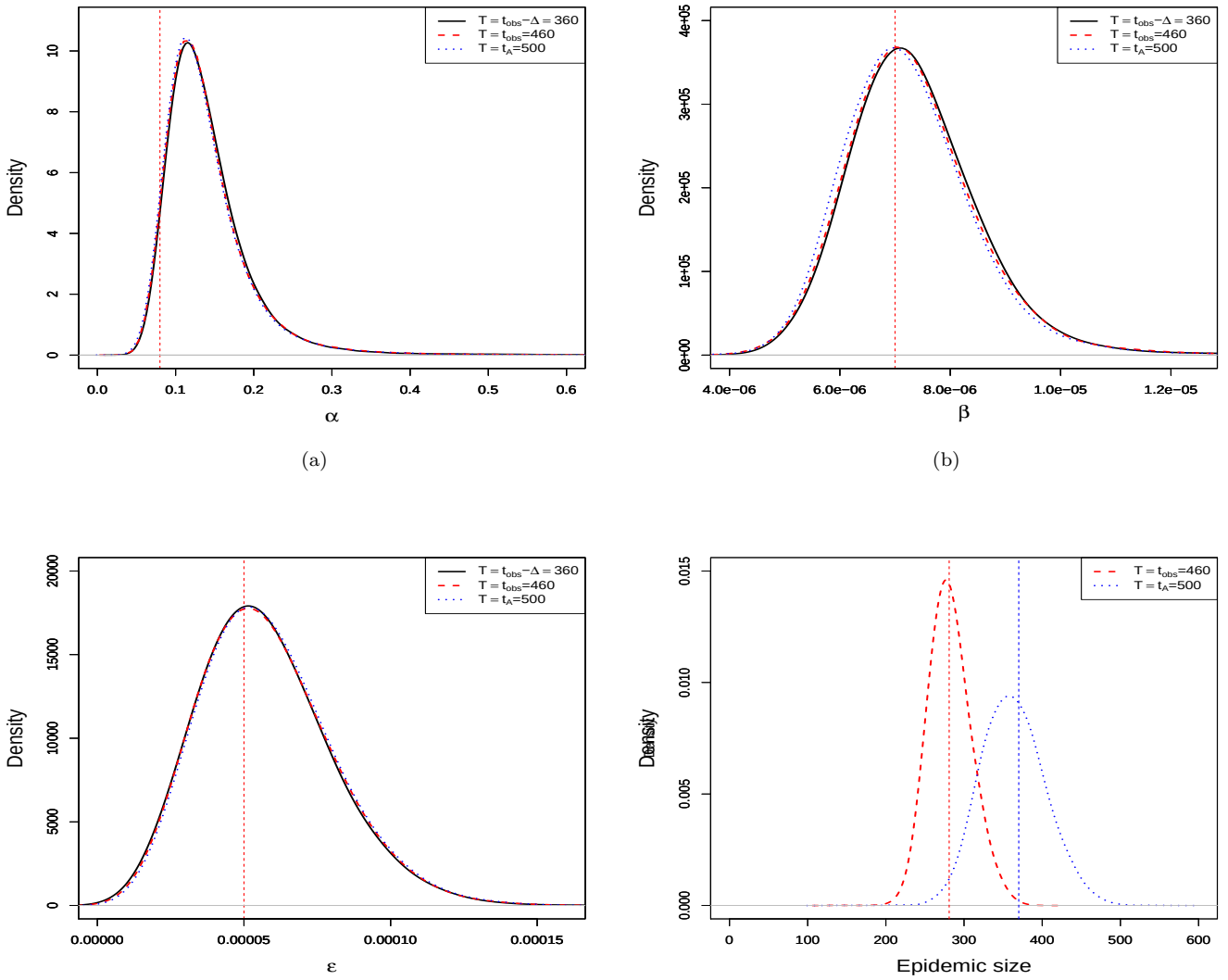
Figure 2: Sample of the disease progress maps made at 30–day intervals from $t_0 = 130$ up to $t_{obs} = 460$, on a population of size $N = 1000$ from simulated data. Symptomatic hosts, cryptic infections and susceptible hosts at the time of the snapshot are denoted by red, blue and white dots respectively.

3.2 Parameter estimation

We implement the MCMC algorithm of Section 1 to sample from the posterior distribution $\pi(\theta, \underline{x}(T)|\mathbf{y})$ for various values of T, namely, $T = t_{obs} - \Delta$, $T = t_{obs}$ and $T = t_A$ where $t_{obs} = 460$, $t_A = 500$ and $\Delta = 100$. These are used to estimate the three measures (risk, hazard and thread) at $t_M = t_{obs}$ and $t_M = t_A$.

We run the algorithms for 280000 iterations for the case $T = 360$ and 10^6 iterations for $T = 460$ and $T = 500$ discarding the first 10000 iterations to ensure that convergence to stationary distribution is reached. The trace plots in Figure 4 show that the chains are mixing well and exhibit no sign of non-convergence. The posterior distributions of the parameters along with the epidemic size at $T = 460$ and $T = 500$ are shown in Figure 3. We can observe from this figure that the true parameter values (dashed lines) are consistent with their respective posterior distributions. The posterior distributions for different T shown on Figure 3 suggest that estimated $\pi_0(\boldsymbol{\theta}|\mathbf{y})$ is the same regardless of which algorithm is used and how far beyond t_{obs} we impute infection times. This provides evidence for the validity of the implementation of the MCMC algorithms.

We use 100000 samples from the joint posterior distribution of the model parameters and hosts infection times to construct the risk, hazard and the threat map as described in the main text at two different times, $t_M = t_{obs} = 460$ and at $t_M = t_A = 500$ (see figure5).



(a) (b)

(c) (d)

Figure 3: The posterior distributions for Bayesian MCMC estimation of the model parameters and the posterior distribution of the epidemic size at T for $T = 460, 500$ days. Dashed lines correspond to the parameter values used for the simulation((a)-(c)) and the simulated epidemic size at $T = 460, 500$ ((d)).

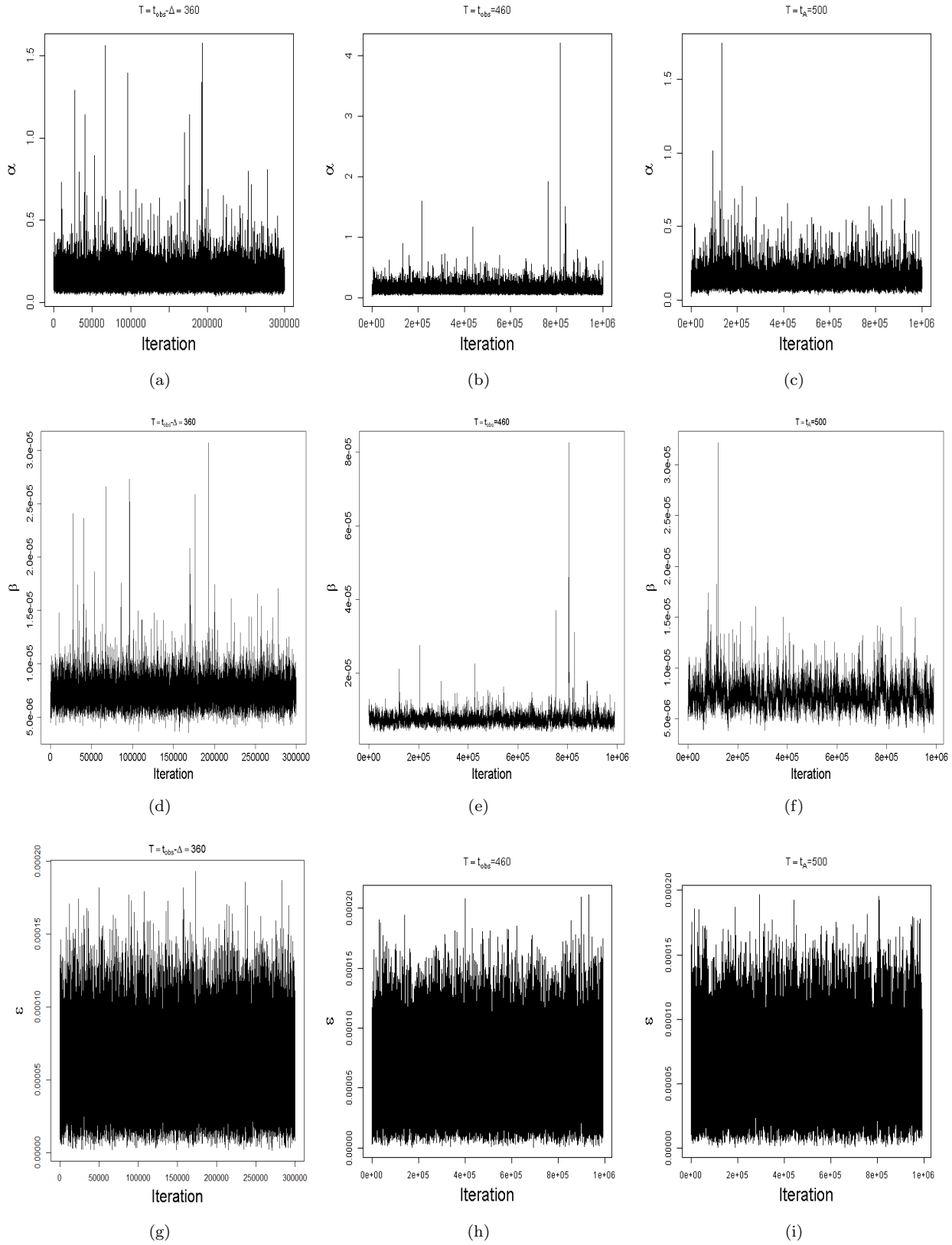


Figure 4: Sample trace plots for α, β and ϵ after a burn-in of 10000 iterations with no augmentation period ($T = 360$) (a), (d) and (g) and with different augmentation periods: $T = 460$ (b), (e) and (h) and $T = 500$ (c), (f) and (i).

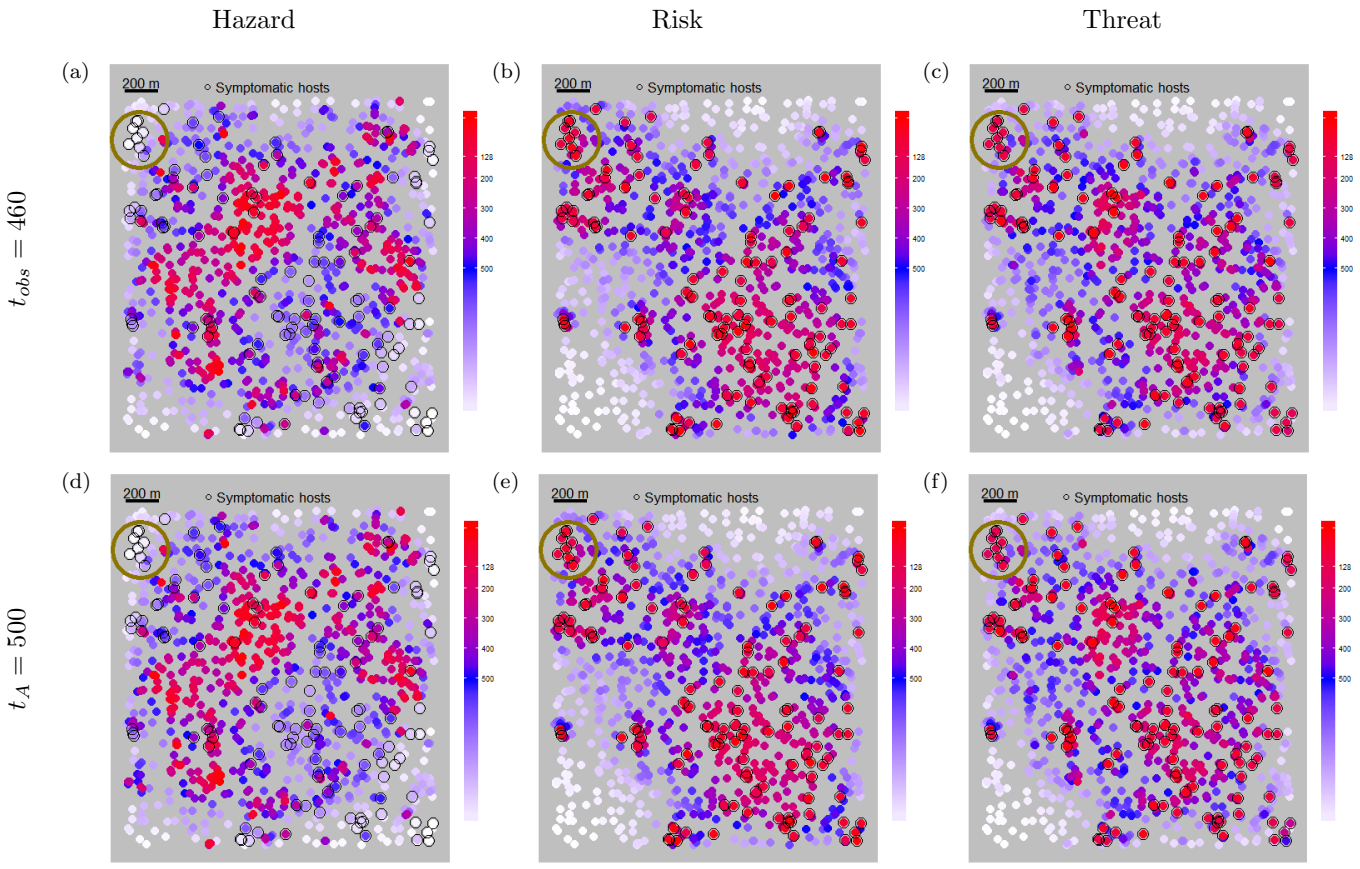


Figure 5: Posterior predictive maps of the threat, risk and threat measures calculated at both $t_M = 460$ ((a) - (c))and $t_M = 500$ ((d) - (f))for the simulated epidemic on the uniformly distributed host population (Section 3.1). Each point represents an individual host with colour varying from white to blue to red with increasing values of the respective measure for that host. The 128 symptomatic hosts detected during the survey are indicated by the black circles. Note that the hosts with high hazard values ((a) and (d)) are located in regions of low infection while the the risk measure is greatest for symptomatic individuals. The dependence of the threat measure on the positions of likely susceptible individuals in relation to an infected host can be discerned. For example, the cluster of infected hosts (circled) in the top left corner of the population naturally exhibit high values of the risk while the corresponding threat measure is comparatively lower, as a high proportion of their immediate neighbours are already infected.

4 Example of clustered hosts population using a short range dispersal kernel

4.1 Model specification

SI models on a population of $N = 1111$ citrus trees from Broward county in Florida labelled B_2 (see figure 6. An epidemic of citrus canker on this population was analysed by Neri et al. (2014). Two different epidemics are generated using a short-range dispersal kernel $K(d, \alpha) = \frac{1}{2\pi d}\frac{1}{\alpha}\exp(-d/\alpha)$: with primary infection (Case (I)) and without primary infection (Case (II)). Parameters used for the simulation are summarized in Table 1.

Case	α	β	ϵ	t_{obs}	Infections observed	Cryptic	T
(I)	0.08	7.10^{-6}	0.00005	460	169	133	500
(II)	0.08	8.10^{-6}	0	460	111	124	500

Table 1: Summary of the parameters used and outcomes obtained for the simulations in the three cases.

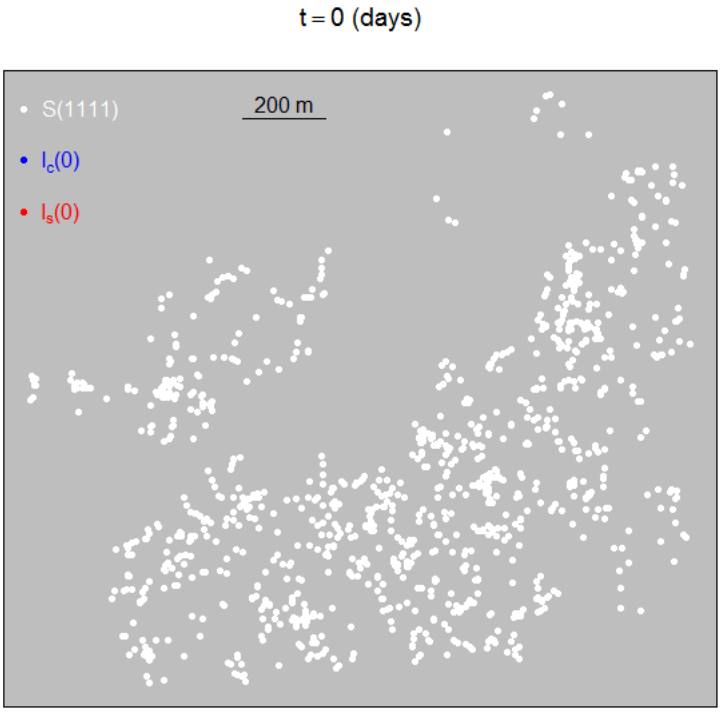


Figure 6: Citrus locations from Broward county.

4.2 Parameter estimation

For the parameter estimation, we adopt the MCMC algorithm described in Section 1. The estimation is done as for the uniformly distributed host population varying T depending on the case considered (see Figures 7, 8). In all cases, the algorithm is run for 520000 steps with a burn-in period corresponding to the initial 20000 iterations.

In Figures 9, 10 we show the sample trace plots of the parameters in all cases. The convergence issues are of no concern as shown on the figures. Nevertheless, it is clearly apparent that the chain mixing depends on how far we augment the imputation period. For instance, the sample trace plot for parameters α in Figures 9a and 9b, and β in 9d and 9e suggest that complete knowledge on the number of infections ($T = t_{obs} - \Delta$) leads to a chain which mixes better than when precise times of future infections are imputed explicitly.

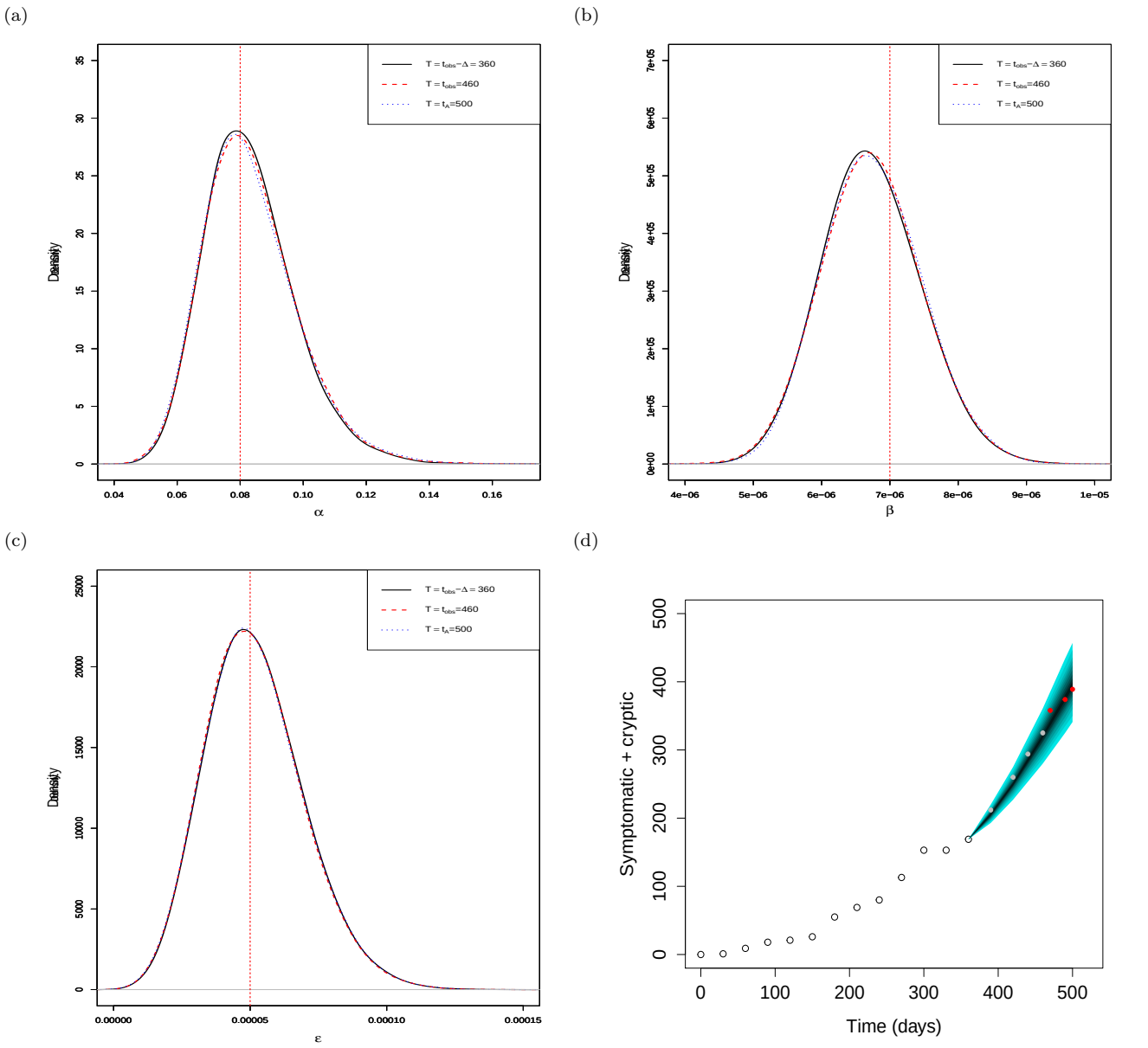


Figure 7: Case (I). Marginal posterior distributions of the model parameters, estimated from MCMC sampling of $\pi(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$, including the dispersal rate α (a), the secondary infection rate β (b), and the primary infection rate ϵ (c), for $T = 360$ (black), $T = 460$ (red) and $T = 500$ (blue). Dashed lines correspond to the parameter values used for the simulation. (d) 95% posterior credible band of the distribution of the disease progress during the period $[360, 500]$ (shaded region) compared to the actual disease progress (open circles represent the trajectory up to $t_{obs} - \Delta$, gray dots represent the trajectory at from $t_{obs} - \Delta$ to t_{obs} and red dots represent the subsequent trajectory up to $t_A = 500$).

The posterior distribution of the model parameters α, β and ϵ for various T for Cases (I) and (II), shown respectively in Figures 7 and 8, match regardless of how far we impute infection times beyond t_{obs}. This gives evidence that the algorithm gives an accurate picture of the posterior distribution. In addition, it can be seen that the model parameters used for the simulation (dashed lines) are consistent with their respective posterior distributions highlighting the fact that the estimation is good. We estimate the epidemic size considering two different augmentation periods including $T = t_{obs}$ and $T = t_A$ as shown in Table 1. Figure 7d shows the 95% credible band of the epidemic trajectory for Case (I). It can be seen that the actual trajectory of the epidemic is contained in the 95% credible regions. Also, the posterior distribution of the predicted epidemic size is shown in Figure 8c for case (II). Again, this suggests that the actual epidemic size lies within the range of the values supported by the predicted distribution.

To prioritise hosts for control, a range of maps are constructed at $t_{obs} = 460$, $t_C = 470$, $t_A = 500$ shown on figures 11-12, using 100000 samples from the joint posterior distribution of the model parameters and hosts infection times. Control results are reported in table 2.

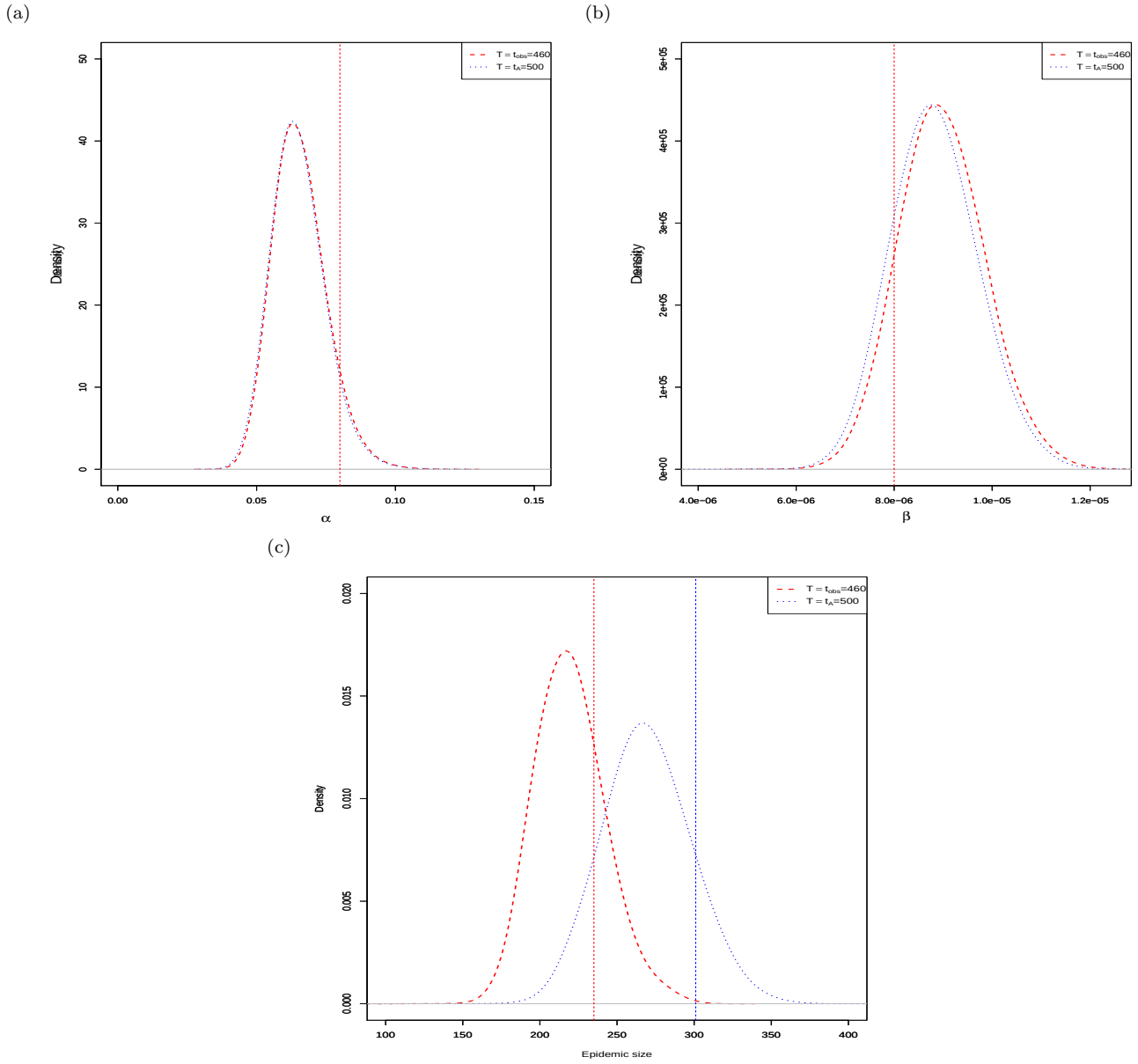


Figure 8: Case (II) The posterior distributions of the model parameters, estimated by sampling from $\pi(\boldsymbol{\theta}, \underline{x}(T)|\mathbf{y})$ using MCMC, including the dispersal rate α (a), the secondary infection rate β (b) using $T = 460$ (red) and $T = 500$ (blue). Vertical lines correspond to the actual parameter value used for the simulation. (c) The posterior distributions of the epidemic size at these times. The simulated values are shown as vertical lines

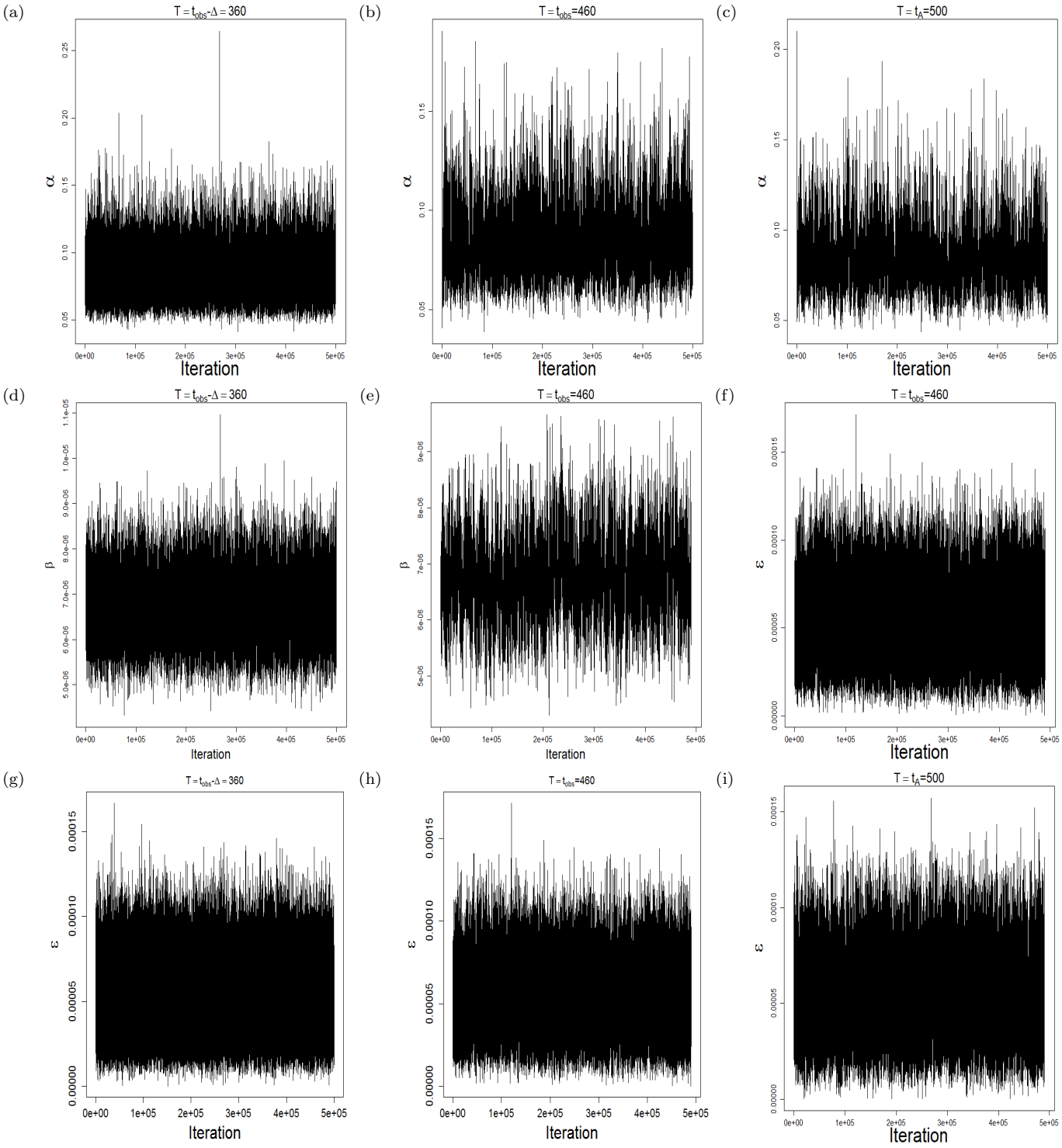


Figure 9: Case (I): Sample trace plots for the posterior distribution of parameters α β and ϵ after a burn-in period of 10000 iterations using the MCMC algorithms described in Section 1.

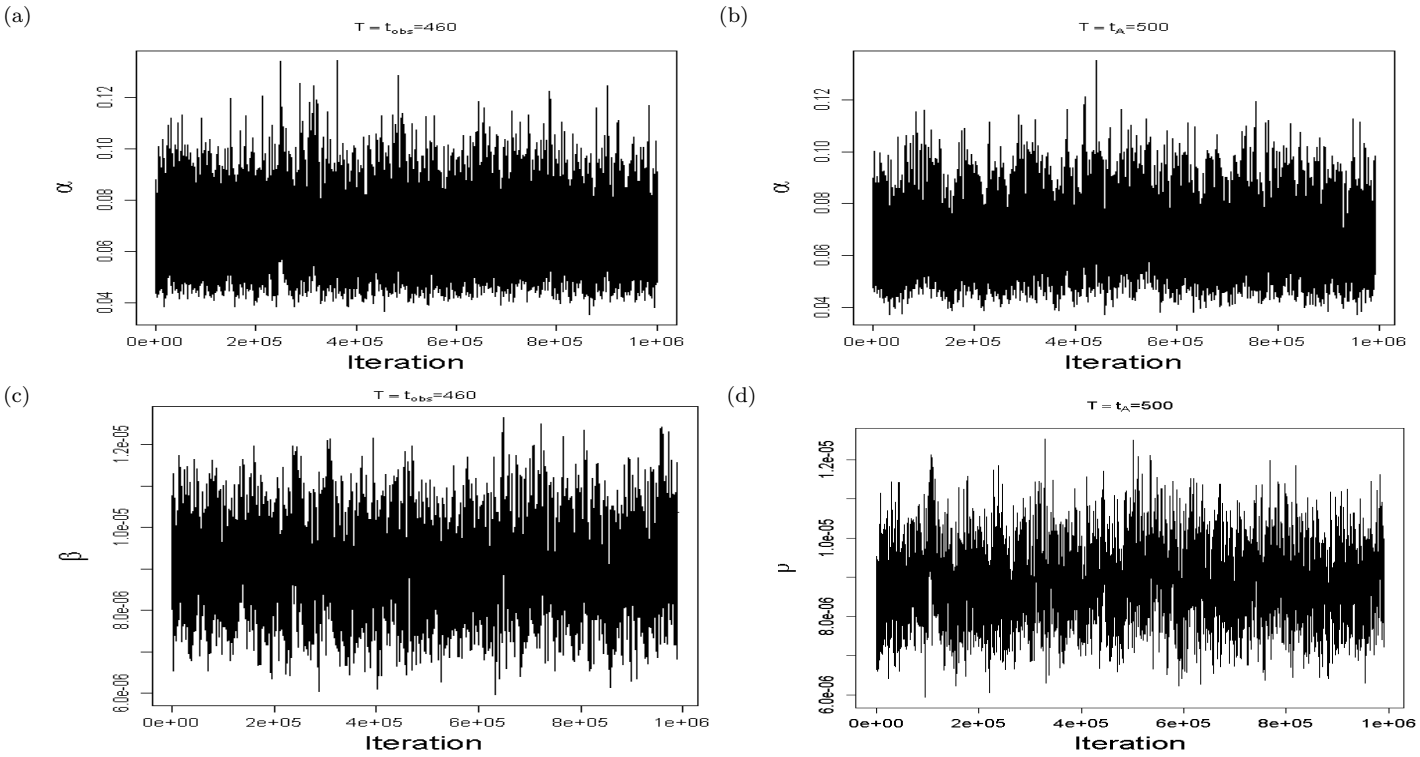


Figure 10: Case (II): Sample trace plots for the posterior distribution of parameters α β after a burn-in period of 10000 iterations using the MCMC algorithms described in Section 1.

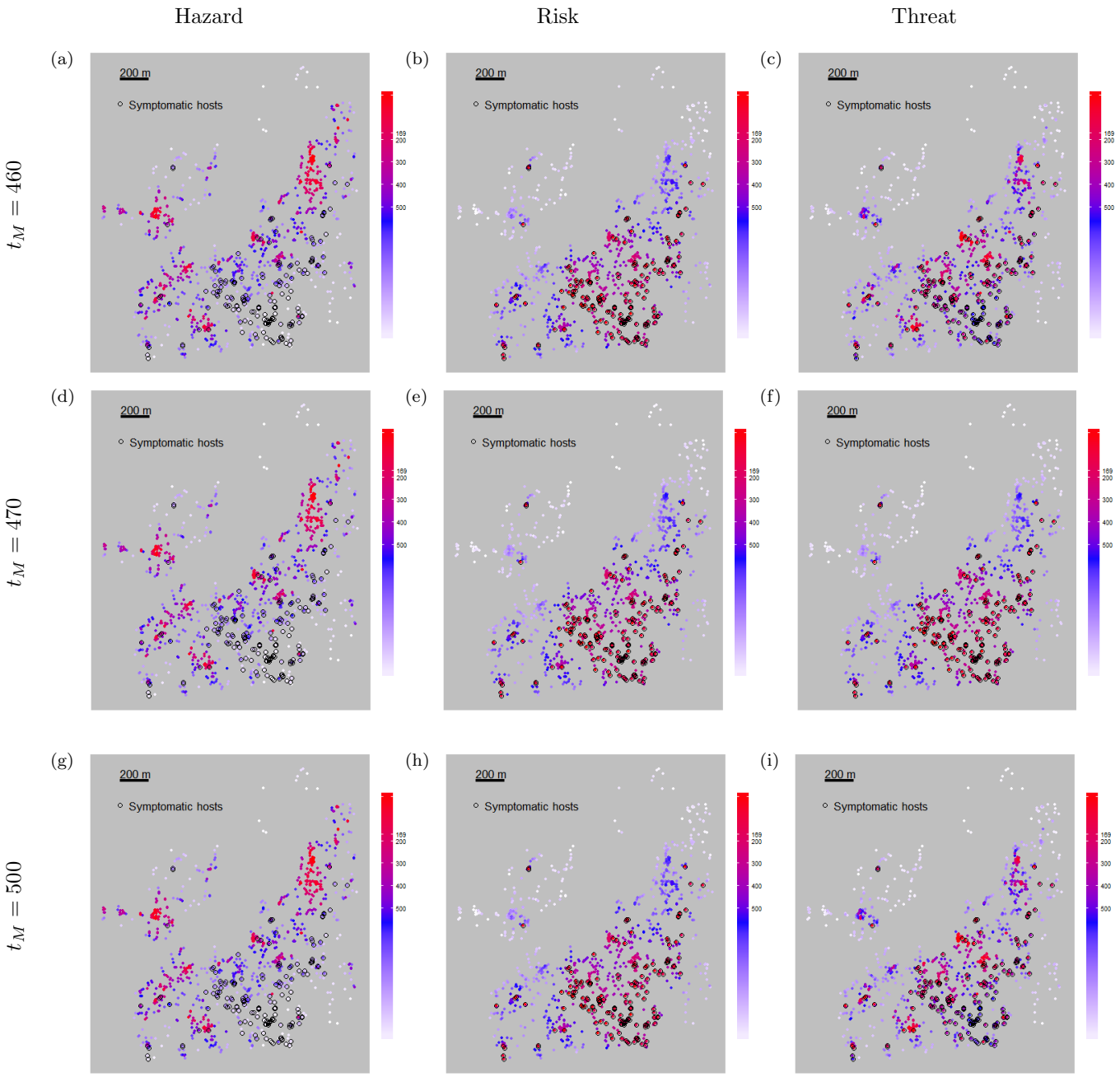


Figure 11: Posterior predictive maps of the risk and threat measures at $t_M = 460$, $t_M = 470$ and $t_M = 500$ for Case (I). The colour of points exhibits a gradation from from white to red with increasing values of the respective measure. The 169 symptomatic hosts detected during the survey are indicated by the black circles.

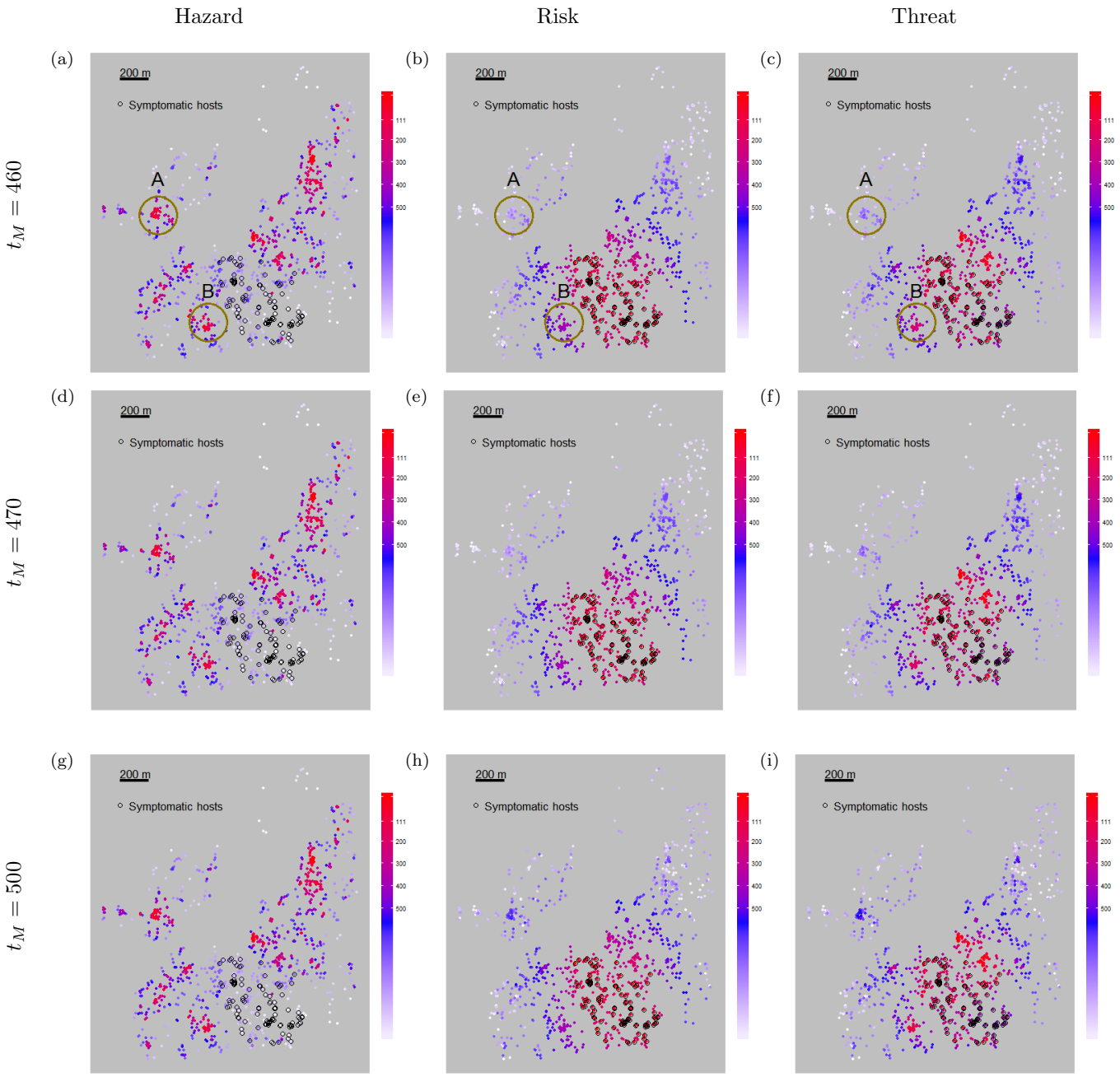


Figure 12: Posterior predictive maps of the risk and threat measures for $t_M = 460$, $t_M = 470$ and $t_M = 500$ for Case (II). The colour of points exhibits a gradation from from white to blue to red with increasing values of the respective measure. The 111 symptomatic hosts detected during the survey are indicated by the black circles.

		Expected number removed			Expected reduction in number infected		
	N'	$\mathcal{H}^*$	$\mathcal{R}^*$	$\mathcal{T}^*$	$\mathcal{H}^*$	$\mathcal{R}^*$	$\mathcal{T}^*$
$t_C = 460$				$t_M = 460$			
				Citrus location: Exponential kernel with primary and secondary infection			
	169	28.011 (0.293)	169 (0)	117.132 (0.31)	17.726 (0.236)	24.034 (0.164)	34.618 (0.217)
	200	33.538 (0.32)	187.475 (0.106)	139.854 (0.333)	20.16 (0.249)	27.886 (0.18)	38.711 (0.234)
	300	52.211 (0.41)	233.067 (0.342)	190.622 (0.424)	28.746 (0.293)	42.312 (0.226)	48.98 (0.29)
	400	72.644 (0.471)	264.845 (0.479)	234.586 (0.503)	36.088 (0.324)	51.91 (0.261)	56.187 (0.318)
	500	89.028 (0.532)	287.201 (0.579)	266.840 (0.567)	40.867 (0.348)	58.308 (0.29)	61.957 (0.356)
				Citrus location: Exponential kernel with no primary infection			
	111	8.419 (0.255)	111 (0)	66.454 (0.366)	9.37 (0.242)	12.82 (0.125)	23.423 (0.271)
	200	16.173 (0.345)	165.325 (0.201)	120.113 (0.473)	15.022 (0.293)	22.232 (0.167)	34.925 (0.314)
	300	21.416 (0.429)	198.759 (0.478)	165.351 (0.571)	18.936 (0.324)	37.89 (0.277)	44.047 (0.367)
	400	26.268 (0.479)	214.429 (0.664)	210.897 (0.662)	21.288 (0.342)	48.77 (0.374)	49.678 (0.39)
	500	31.819 (0.529)	218.612 (0.706)	218.660 (0.707)	23.966 (0.359)	51.931 (0.403)	52.132 (0.407)
				$t_M = 500$			
				Citrus location: Exponential kernel with primary and secondary infection			
	169	21.077 (0.253)	169 (0)	95.159 (0.344)	14.249 (0.229)	24.034 (0.164)	33.602 (0.237)
	200	27.022 (0.288)	187.238 (0.136)	115.930 (0.361)	17.470 (0.245)	29.432 (0.186)	37.883 (0.259)
	300	47.402 (0.388)	232.587 (0.354)	166.497 (0.434)	26.876 (0.289)	43.183 (0.231)	48.406 (0.306)
	400	67.656 (0.454)	264.341 (0.481)	202.066 (0.5)	34.486 (0.32)	52.609 (0.266)	55.618 (0.342)
	500	81.683 (0.51)	287.285 (0.579)	237.492 (0.566)	38.775 (0.343)	58.634 (0.29)	61.469 (0.371)
				Citrus location: Exponential kernel with no primary infection			
	111	6.555 (0.231)	111 (0)	47.421 (0.428)	7.753 (0.231)	12.82 (0.125)	23.442 (0.298)
	200	10.692 (0.297)	165.402 (0.204)	95.946 (0.507)	11.833 (0.279)	22.963 (0.171)	34.583 (0.34)
	300	17.098 (0.371)	199.067 (0.481)	145.832 (0.597)	16.458 (0.309)	37.661 (0.266)	43.926 (0.372)
	400	23.827 (0.462)	214.453 (0.666)	192.642 (0.669)	20.444 (0.335)	48.565 (0.373)	49.539 (0.394)
	500	27.758 (0.499)	219.120 (0.709)	218.636 (0.708)	22.126 (0.351)	52.053 (0.405)	52.194 (0.408)
$t_C = 470$				$t_M = 470$			
				Citrus location: Exponential kernel with primary and secondary infection			
	169	30.643 (0.312)	169 (0)	116.688 (0.332)	14.162 (0.197)	17.616 (0.14)	26.455 (0.187)
	200	35.587 (0.341)	188.849 (0.115)	138.736 (0.351)	15.853 (0.209)	21.238 (0.152)	29.643 (0.205)
	300	57.377 (0.44)	238.692 (0.351)	190.458 (0.45)	22.774 (0.242)	32.188 (0.192)	37.642 (0.245)
	400	78.507 (0.512)	273.329 (0.491)	235.038 (0.529)	28.18 (0.268)	39.277 (0.219)	43.150 (0.27)
	500	96.699 (0.575)	298.318 (0.605)	270.113 (0.606)	32.148 (0.285)	44.292 (0.24)	47.714 (0.3)
				Citrus location: Exponential kernel with no primary infection			
	111	9.348 (0.281)	111 (0)	66.392 (0.425)	7.552 (0.2)	9.328 (0.106)	18.886 (0.22)
	200	17.356 (0.382)	168.756 (0.195)	118.735 (0.533)	12.556 (0.24)	16.383 (0.137	27.524 (0.267)
	300	24.825 (0.47)	205.676 (0.505)	168.998 (0.628)	15.749 (0.267)	29.332 (0.2366)	34.294 (0.295)
	400	31.065 (0.537)	224.098 (0.714	215.777 (0.721)	18.002 (0.281)	37.738 (0.305)	38.815 (0.316)
	500	36.038 (0.585)	229.6 (0.767)	229.664 (0.77)	19.725 (0.293)	40.448 (0.324)	40.826 (0.33)
				$t_M = 500$			
				Citrus location: Exponential kernel with primary and secondary infection			
	169	24.512 (0.284)	169 (0)	100.071 (0.358)	11.901 (0.19)	17.616 (0.14)	25.84 (0.197)
	200	31.120 (0.321)	188.691 (0.134)	121.277 (0.377)	14.461 (0.203)	21.838 (0.155)	29.194 (0.21)
	300	53.554 (0.427)	238.316 (0.356)	173.924 (0.458)	21.743 (0.242)	32.507 (0.193)	37.308 (0.252)
	400	75.237 (0.4981)	273.415 (0.496)	211.419 (0.538)	27.39 (0.265)	39.801 (0.222)	42.983 (0.288)
	500	90.647 (0.561)	298.987 (0.605)	248.874 (0.608)	30.918 (0.285)	44.694 (0.241)	47.449 (0.309)
				Citrus location: Exponential kernel with no primary infection			
	111	8.27 (0.267)	111 (0)	51.55 (0.47)	6.911 (0.194)	9.328 (0.106)	18.965 (0.247)
	200	13.198 (0.342)	168.664 (0.198)	102.101 (0.557)	10.337 (0.235)	16.89 (0.142)	27.351 (0.274)
	300	20.599 (0.421)	206.42 (0.502)	154.165 (0.65)	14.137 (0.26)	28.849 (0.221)	34.448 (0.302)
	400	28.387 (0.52)	224.545 (0.718)	202.693 (0.726)	17.366 (0.278)	37.769 (0.304)	38.832 (0.319)
	500	32.799 (0.56)	230.169 (0.772)	229.645 (0.772)	18.693 (0.288)	40.759 (0.328)	40.858 (0.33)

Table 2: Estimates of expected number removed and expected reduction in number infected by $t_A = 500$ with the standard error obtained using various values of N' where the controls are applied at $t_C = 460$ and $t_C = 470$.

5 Sensitivity to dispersal kernel parameter

We tested the effect of varying the dispersal scale around the default value (case I- see table 1). We repeat the same analysis considering four values of the dispersal scale (see table 3). In all cases, we simulate the epidemics (using Sellke algorithm) with the assumption that the first infection occurs at the same location as in the previous cases.

Here, we only consider the estimation at $t_M = t_A = 500$ and the MCMC is run for 100000 iterations. Figure 13 shows that the true values of the parameters lie in a 95% credible interval of their posterior distributions. Also, the trace plots (not shown here) do not show evidence of non convergence - evidence that the estimation performs well.

Case	α	β	ϵ	t_{obs}	Infections observed	Cryptic	t_A
Simulation 1	0.015	7.10^{-6}	0.00005	460	148	100	500
Simulation 2	0.04	7.10^{-6}	0.00005	460	169	84	500
Simulation 3	0.16	7.10^{-6}	0.00005	460	121	146	500
Simulation 4	0.2	7.10^{-6}	0.00005	460	100	106	500

Table 3: Summary of the parameters used and outcomes obtained for the simulations in the three cases.

Similarly, we implement the control strategy described in the main text and show the effect of varying N', the number of hosts consider for removal, on the estimated values of the expected infections, expected reduction (with respect to the no-control) and the expected number of removals (see figures 14-17). Results show that as the number of hosts to remove increases, control based on the risk map becomes as effective as the one based on the threat map. However, results of Simulation 3 and 4 (see figure 16 and 17) show that removal based on the threat and the risk maps are qualitatively similar. This may be due to the fact that for larger values of the dispersal scale parameter, long-distance host-to-host infection occurs. Therefore, the imputed values of the hazard do not exhibit as much variation over hosts, leading to the threat measure being dominated by the risk measure.

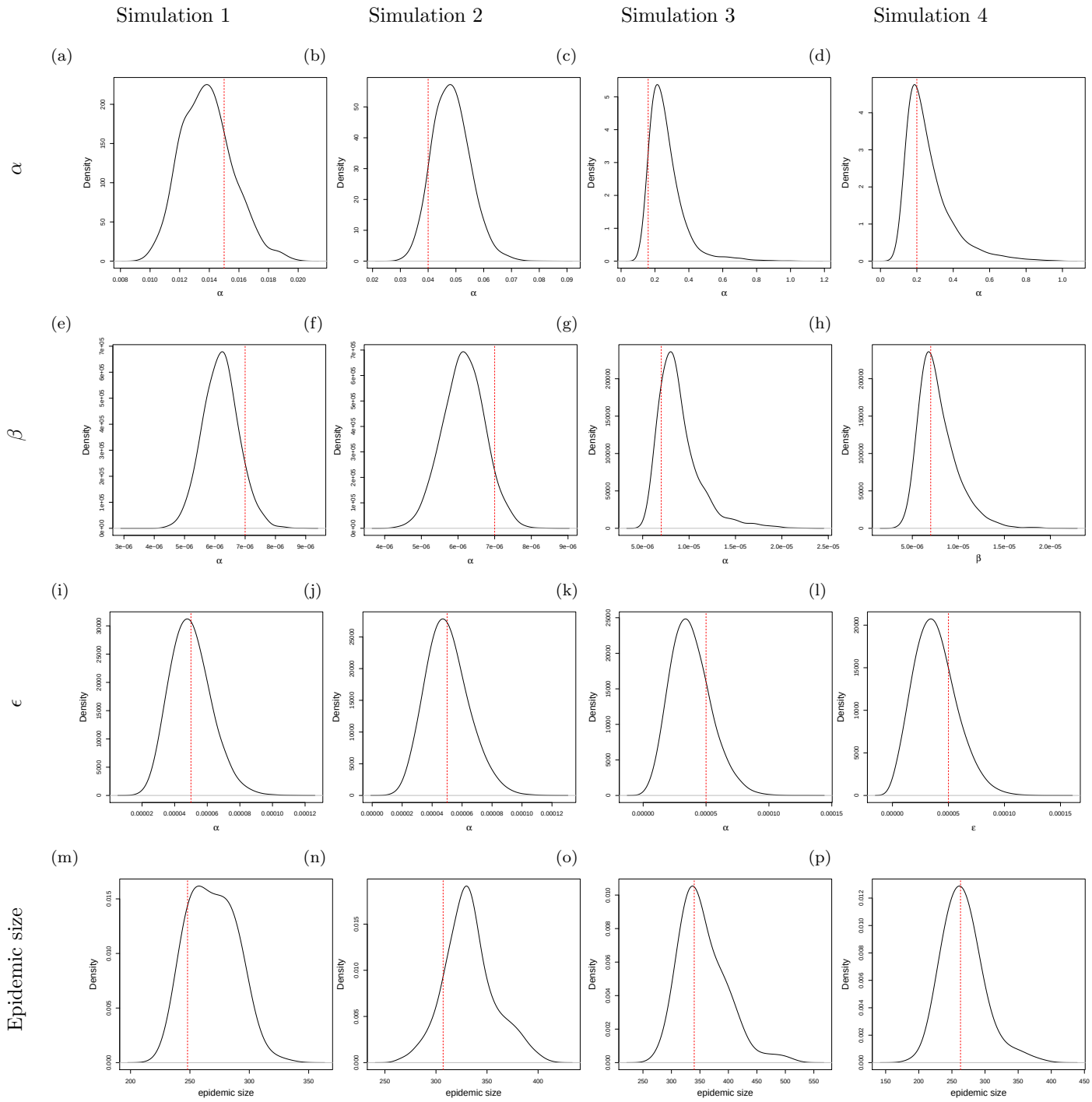


Figure 13: Posterior distribution of the model parameters estimated at $t_A = 500$. The red line indicate the true values.

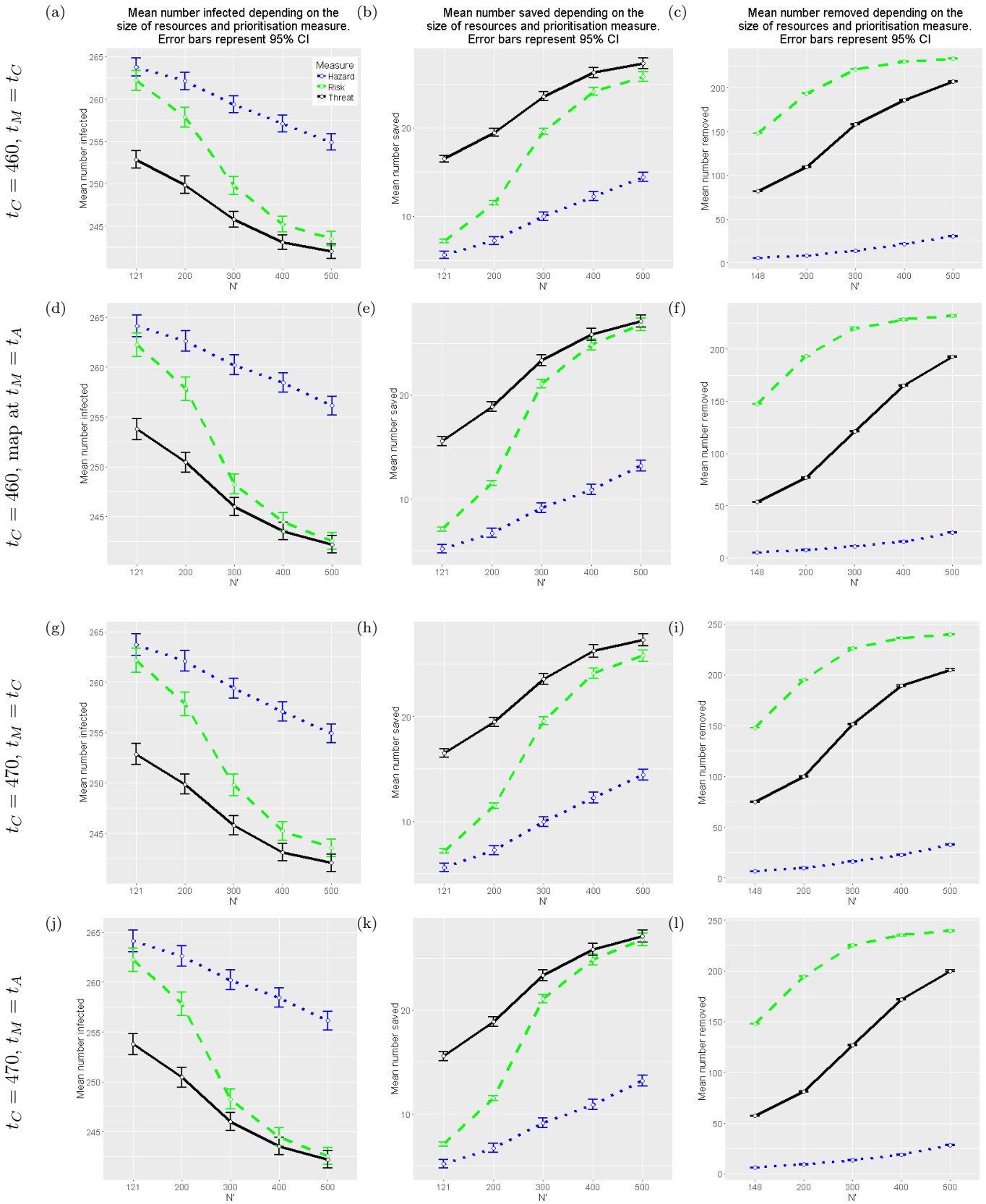


Figure 14: **Simulation 1**: Marginal confidence intervals for the exected number of infections by $t_A = 500$, the estimated expected reduction in infection with respect to the no-control case, and the expected number of removals using a exopnential kernel (for $\alpha = 0.015$).

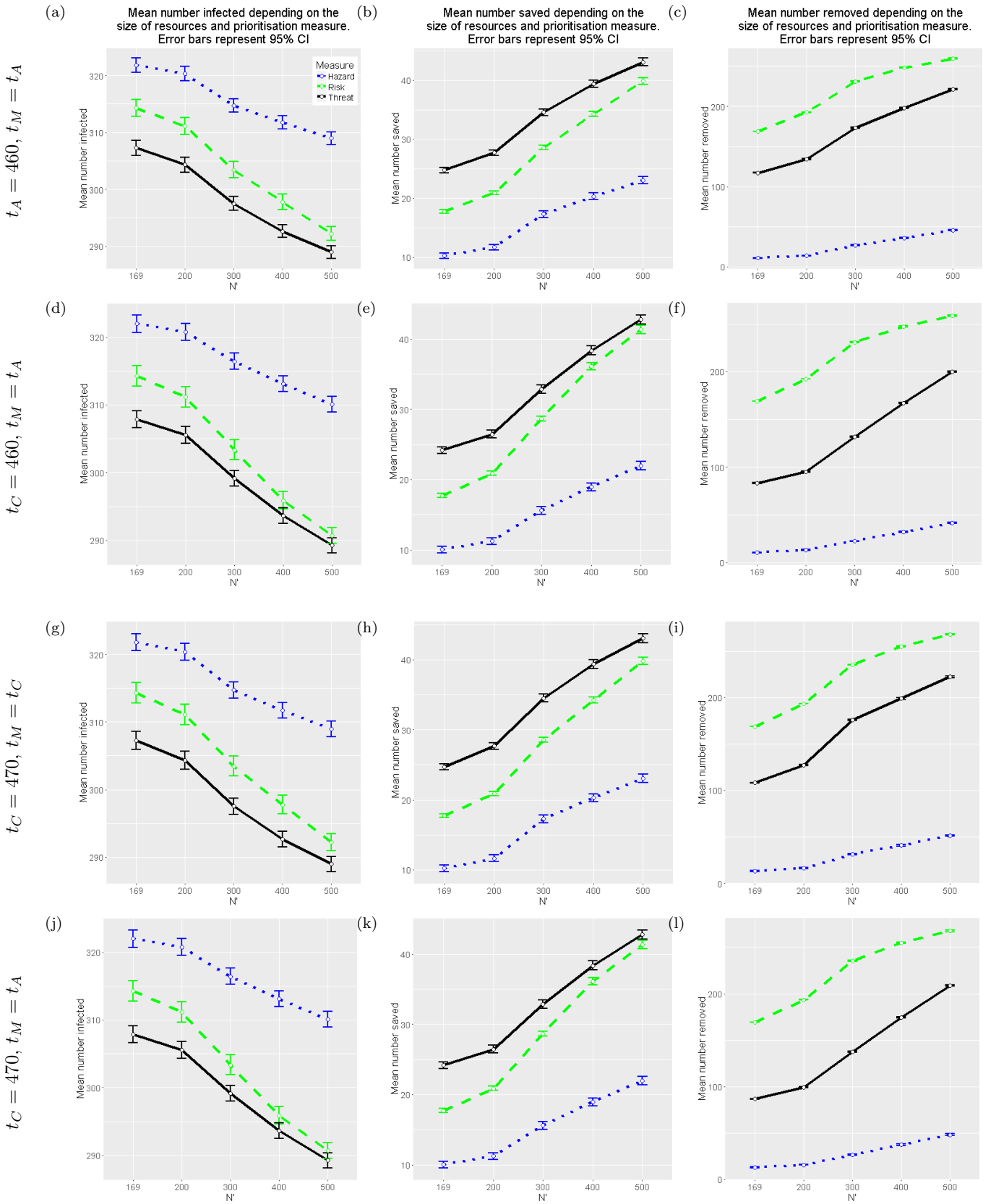


Figure 15: **Simulation 2**: Marginal confidence intervals for the exected number of infections by $t_A = 500$, the estimated expected reduction in infection with respect to the no-control case, and the expected number of removals using a exopnential kernel (for $\alpha = 0.04$).

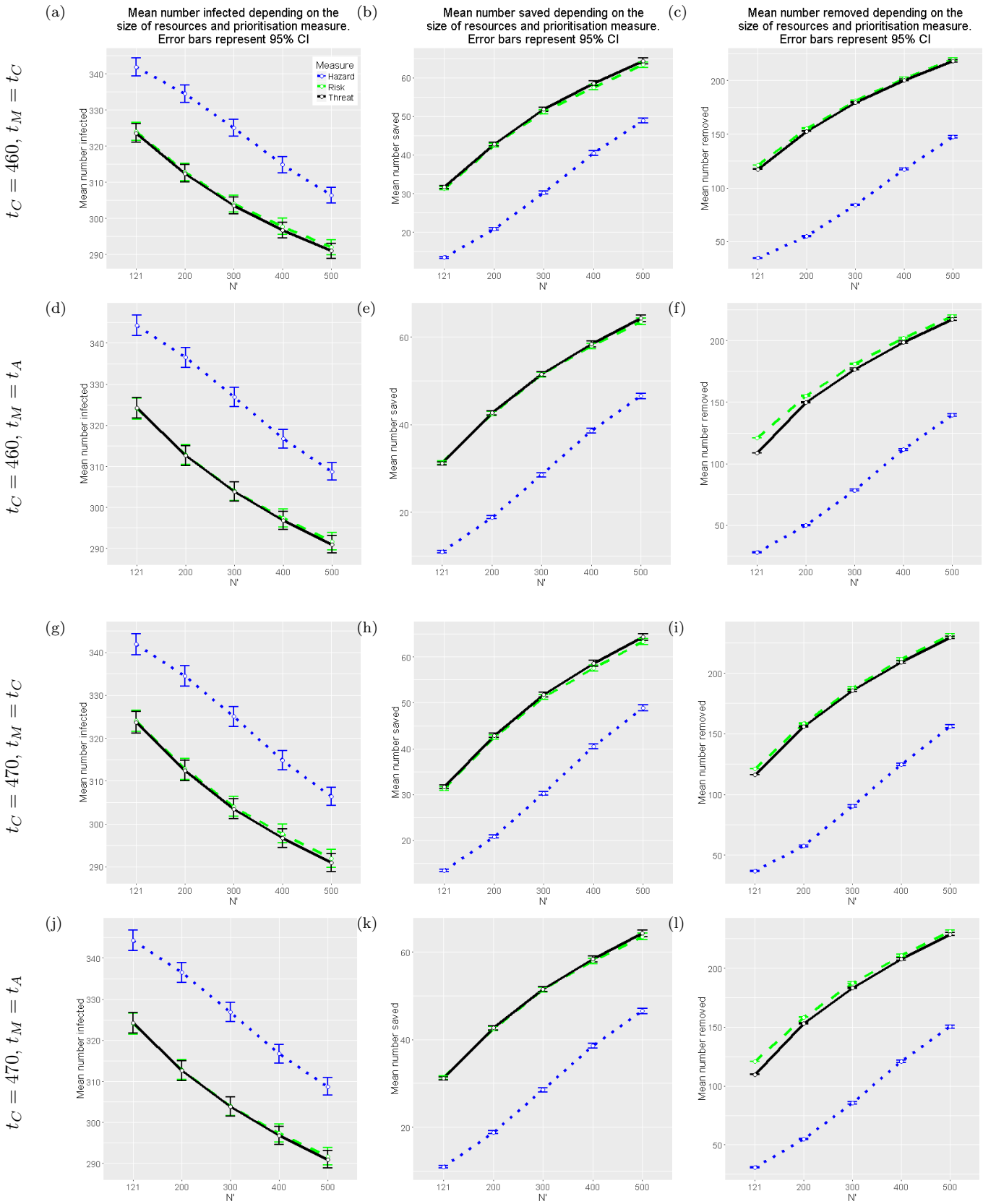


Figure 16: **Simulation 3**: Marginal confidence intervals for the exected number of infections by $t_A = 500$, the estimated expected reduction in infection with respect to the no-control case, and the expected number of removals using a exopnential kernel (for $\alpha = 0.16$).

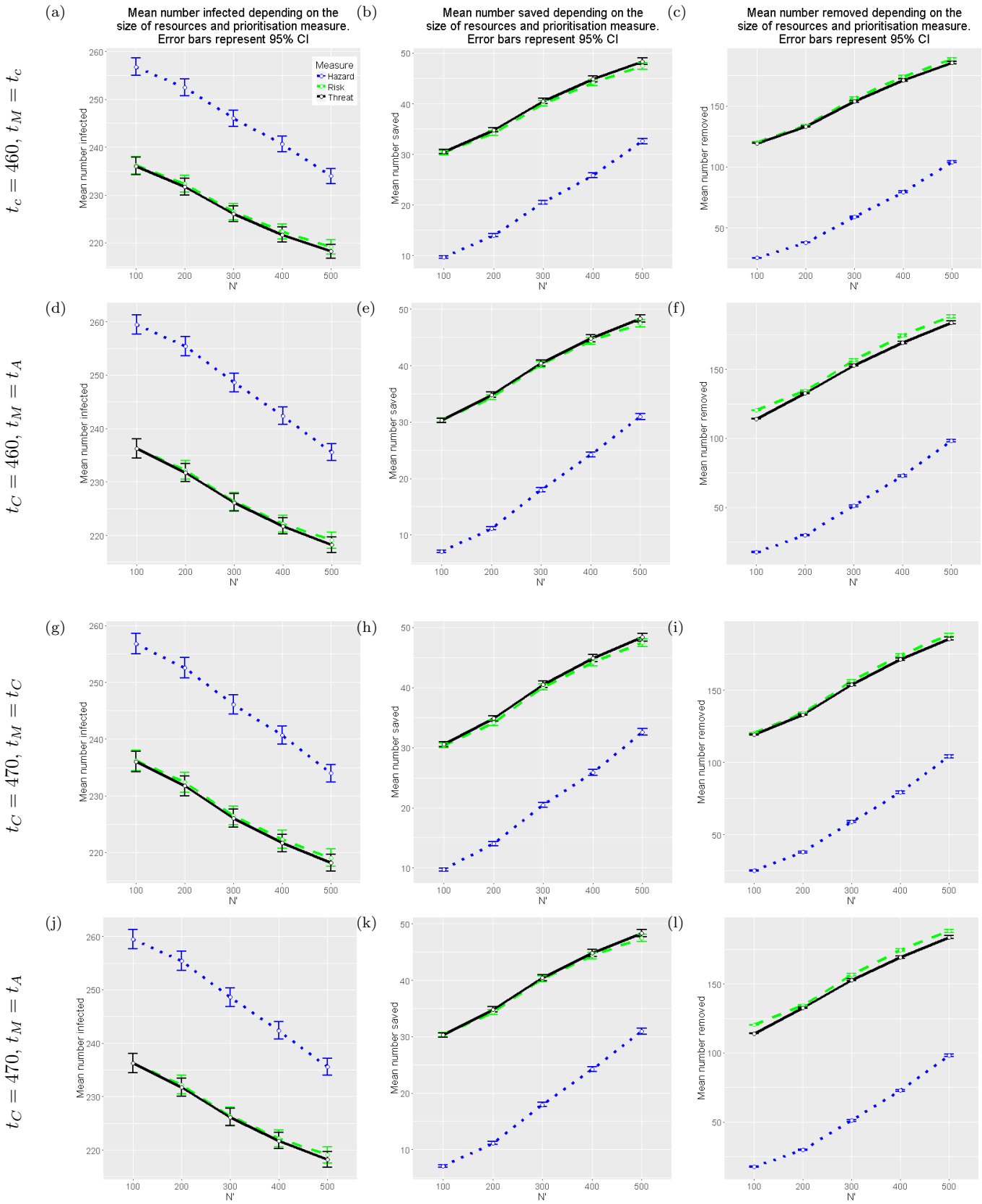


Figure 17: **Simulation 4**: Marginal confidence intervals for the exected number of infections by $t_A = 500$, the estimated expected reduction in infection with respect to the no-control case, and the expected number of removals using a exopnential kernel (for $\alpha = 0.2$).

References

Gibson, G J and Renshaw, E. (1998). Estimating parameters in stochastic compartemental model using Markov chain methods. *IMA J Math App Med*, 15:19–40.

Gottwald TR, Sun X, Ripley T, Graham JH, Ferrandino F, Taylor EL (2001) Geo-Reference Spatiotemporal Analysis of the Urban Canker Epidemic in Florida. *Phytopathology*, 92:361–377.

Lau, MSY, Marion, G, Streftaris, G, and Gibson, GJ. (2015). A systematic bayesian integration of epidemiological and genetic data. *PLoS Comput Biol*, 11:e1004633.

Neri FM, Cook AR, Gibson GJ, Gottwald T, Gilligan CA (2014) Bayesian analysis for inference of an emerging epidemic: Citrus canker in urban landscapes. *PLoS Comput Biol*, 10):e1003587.

O'Neill PD, Roberts GO (1999) Bayesian inference for partially observed stochastic epidemics. *J R Stat Soc A*, 162:121–129.

Parry M, Gibson GJ, Parnell S, Gottwald TR, Irey MS, Gast TC, Gilligan CA (2014) Bayesian inference for an emerging arboreal epidemic in the presence of control. *PNAS*, 111:6258–6262.

Sellke T (1983) On the asymptotic distribution of the size of a stochastic epidemic. *J Appl Probab*, 20:390–394.

Streftaris G, Gibson G (2004) Bayesian inference for stochastic epidemics in closed populations. *Statist Model*, 4:63–75.