

# Spatio-temporal model for Plague dynamics

Bård Øyvind Kvaal



Master thesis for Modelling and Data Analysis

The Faculty of Mathematics and Natural Sciences  
UNIVERSITY OF OSLO

May, 2011



# Content

1	Introduction	4
2	Data	5
	General description of the data	5
	Pre-processing of the data	5
	Climate data	8
	Missing data	8
3	Method	9
	Bayesian hierarchical model	9
	Model	9
	Inference	12
	MCMC	12
	Model evaluation by DIC	12
4	BUGS	14
	GeoBUGS	16
	Trap	16
	Achieving DIC	17
5	Results	18
	The models	18
	Reporting results	18
	Occupancy	19
	DIC	20
	Model 5	23
	Model 0a and Model 0b	21
	Model 2 - the best model according to the DIC	24
	$\beta_1$ and $\sigma_E$	26
6	Summary / Discussion	27
	Appendix	i
A	BUGS models	ii
B	Trace plots	viii
C	Model 0a	x
D	Adjustment for t=1 in AR(1) process	xi

# 1 Introduction

For years, biologists at CEES, Centre for Ecological and Evolutionary Synthesis, have analyzed sylvatic plague (*Yersinia Pestis*) in an attempt to increase the knowledge of the biological process. One of the datasets analyzed was collected in Kazakhstan for surveillance purposes. This dataset gives unique opportunities for exploring the biological dynamics of sylvatic plague.

The dataset collected in the desert Saryyesik-Atyrau, southeast of lake Balkhash. A large field was divided into squares. Every year, from 1949 to 1995, some squares were tested for plague among great gerbils (*Rhombomys opimus*), which is the main host for plague at this site. The main vector, fleas (*Xenopsylla*), was also tested for plague infection and counted.

The latest analysis on this was done by Lise Heier et al<sup>1</sup>. She made a very specific analysis of the structure. In further analyses, she found no effects from climatic variables.

Others analyzing the Kazakh dataset includes Nils Chr. Stenseth et al<sup>2</sup> and Noelle I. Samia et al<sup>3</sup>. Both Stenseth and Samia used a threshold model.

The Methods used by Stenseth et al. (2006) reduced the above threshold sample size to 120 excluding flea data and 54 when fleas were included. They found climatic influence on plague, but he had no spatial structure as Lise found to be strongly indicated four years later.

The factors assumed by biologists to be the most important for explaining the biological process of sylvatic plague are climate and the amount of rodents present. I therefor investigate these relations. With the analysis of Lise Heier in mind, a spatial structure is included, but is not tested or even visualized in the results. My model is not directly comparable to the model used by Lise Heier. Her focus was on the movement of the plague, and mine is on the factors explaining the process. Also, our methods are quite different.

Through a Bayesian hierarchical approach, I will be able to use more of the information in the dataset than others have been able to through frequentistic methods. Using such a method, I will also take into account the uncertainty in relation to the missing data and observation errors, which often is ignored.

---

<sup>1</sup>Heier et al. (2011) in References

<sup>2</sup>Stenseth et al. (2006) in References

<sup>3</sup>Samia et al. (2007) in References

## 2 Data

### General description of the data

The Kazakh data which this article focuses on, were collected for surveillance purposes from 1949 to 1995. The area were divided into  $(10 \text{ by } 10 \text{ km}^2)$  squares. Every year squares were investigated both in the spring and in the autumn. When I first started to do analyzes, I only used autumn which I was told might give the best data. I intended to include the spring data when calculations were stable, but my analyzes were too computational demanding on the computer to include the spring data. Every comment on data is on the data collected in the autumn.

Every year  $t$ , sites  $\{i, i \in \mathcal{D}_t^p\}$  were chosen to test great gerbils for plague. At these sites  $n_{i,t}$  rodents were collected by traps along a line. The rodents were tested for plague with bacteriological tests,  $y_{i,t}$  rodents tested positive. There were also done serological tests, but I choose not to focus on these.

As mentioned in the introduction, the amount of rodents is assumed to be one of the most important factors for the process of plague. It has been discussed whether one should use the density or the occupancy which is the relative amount of burrows occupied by rodents. The density of the burrows is very closely related to the distance of interactions between rodents. These interactions are what the biologists assume to be important for the spread of the bacteria. By this argument, the discussion concluded that the occupancy is a natural measure to use.

Occupancy were also closely monitored. Every year  $t$  at random sites  $\{j, j \in \mathcal{D}_t^\phi\}$ ,  $m_{j,t}$  burrows were checked for signs of resent activity. The relative amount of burrows occupied was recorded as estimates of the occupancy. This was done for primary squares  $(10 \times 10\text{km}^2)$  for parts of the spatiotemporal field and elsewhere for secondary squares  $(20 \times 20\text{km}^2)$  which consists of  $2 \times 2$  primary squares. The sites in which occupancy is monitored are not necessarily the same as those of plague.

There is some uncertainty for the occupancy data. One are not sure of the routines used to collect these data. It is indicated that  $m_{j,t} = 30$ , but looking at the data, it seems this is not always correct. Also, one are not sure if the  $m_{j,t}$  tested burrows are uncorrelated. There might be some clustering of the rodents which may give correlated samples of inhabited burrows.

Some flea data are available as well, which also have missing data. Biologists are interested in testing the effects of the amount of fleas relative to rodents. However, I will not include this in this model. One problem is that the sampling process is complicated and not simple to model. It is an even bigger problem that including flea data would need much more complicated computations, since the model I use is very computational demanding even without including the flea data.

### Pre-processing of the data

Below there is a map showing the amount of data collected for each primary square over time (Figure 1). The figure reveals that there has been few observations in parts of the investigated field, especially in the outer parts, e.g. in the northwest. Since the occupancy data are on two scales, both primary

and secondary squares, for different areas at different times, a similar map for occupancy would either be two maps, one for primary and one for secondary squares, or a combination of these.

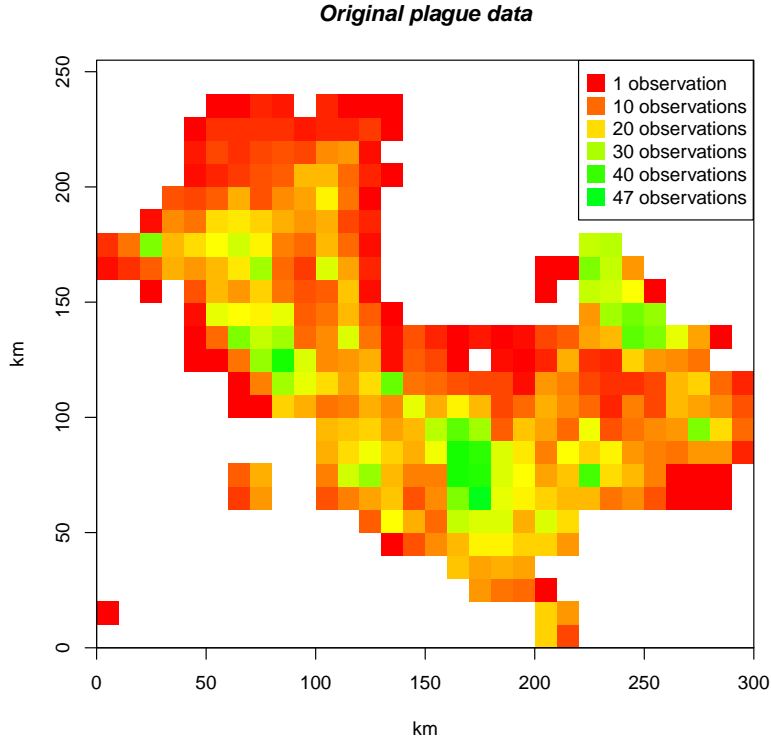


Figure 1: This map shows the amount of collected plague data over time for each primary square. There are several primary squares with only one observation.

I chose to do my analyzes on the secondary squares for two reasons. The first reason is that it is a simple way of using occupancy data both for primary and secondary squares. The second reason is that I reduce the problem with missing plague data by combining four primary squares. The new plague data,  $(y_{i,t}, n_{i,t}, i \in \{1, 2, \dots, N\}, N = \text{the number of secondary squares})$ , for secondary squares is defined as the sums of data for primary squares in each secondary square. The occupancy data is either data collected for secondary squares or the mean of observed data for primary squares in the given secondary square.

To minimize the problems with missing data even more, I used only parts of the dataset. To include a secondary square, I demanded a certain amount of plague data. Every secondary square included should have data for at least 20% of the time points. Included secondary squares should also have at least two included neighboring squares.

Also, looking at the amount of occupancy data over time (Figure 2), I found the first nine years to have too little occupancy data. I decided to exclude these years (1949-1957).

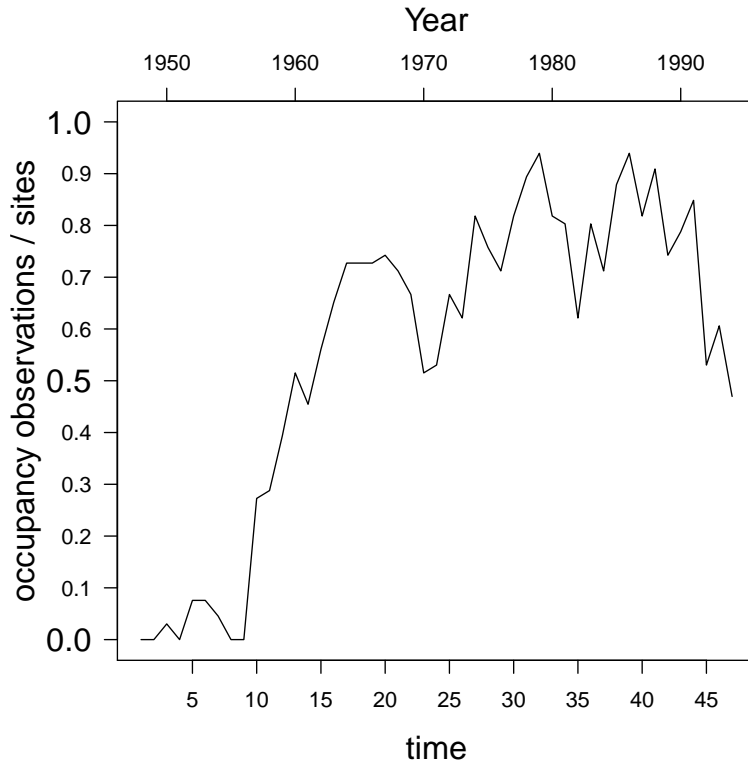


Figure 2: The figure shows how many of the 66 secondary squares that has occupancy data for each time point.

Selecting the data in such a way, ensures a data quality that gives better inference. Every square included in the analysis now gives more information, both through structure and data for every spatiotemporal area, and less noise.

The final data used, consists of 66 secondary squares and 38 years (1958-1995). Figure 3 shows the amount of data used for each square over time. It is easily seen that there is much data for occupancy. Also, comparing this figure with figure 1, we see great improvement of the plague data.

I transformed the occupancy data from the relative amount of burrows inhabited in  $(0, 1)$  to the approximated number of occupied burrows. This is done to make a reasonable observation model. As mentioned, this approximation, which assumes  $m_{j,t} = 30$ , seem to not always be correct. I wil comment more on this in the model description.

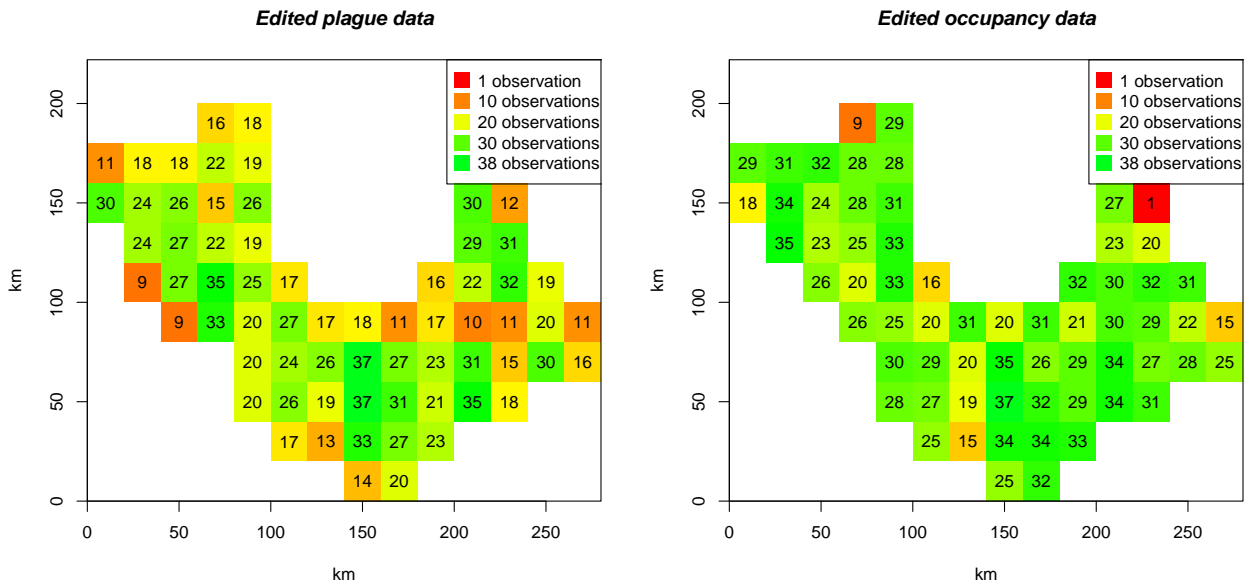


Figure 3: This figure shows the amount of data used for each secondary square. For plague, there is data recorded for at least nine years, which is 23.6% of the time period. For occupancy, there is only two secondary squares with data from less than 15 years.

### Climate data

There have been several indications on climatic influences on plague. Average monthly temperature and average humidity data are collected from a weather station not far from the area in which the occupancy and plague data are collected. I will use these data to search for such relations. The data are standardized for computational reasons.

### Missing data

As mentioned and shown on graphics, there are missing data. In general, this is not a big problem since I use a Bayesian hierarchical model (See section 3). The sites to be tested for plague at a given time were chosen randomly in general, but if there were made reports of sick rodents in a given site, this site would be investigated. I do not know when or how many times such reports were made. This unknown bias in which sites were tested, could change the interpretation of this analysis. This problem has been ignored by most. It has been indicated that there has been few such report. If this is the case, it will not be a very big problem. However, this should be mentioned.



### 3 Method

#### Bayesian hierarchical model

To separate the biological process from the observation process, I use a hierarchical model. I also use a fully Bayesian model, which gives good computational routines even with missing data. The model is now divided in an observation model and an underlying process. In such a way the model handles observation errors which is often ignored, especially for occupancy. The underlying process explains the biology and is what I want to analyze. By using a good structure on this process, one does not need data at every spatiotemporal site. Figure 4 shows that the model is a full Bayesian hierarchical model.

#### Model

The plague data is, as mentioned in the data description, given as  $(y_{i,t}, n_{i,t})$  which might indicate a binomial model for observations. There are three reasons for not doing this. The first reason is that I want to link the analysis to the earlier analysis done by Lise Heier and others. Since she used a binary model, I also did this. Second, there is much uncertainty on how these data are collected. So much so that it would be difficult to model the  $(y_{i,t}, n_{i,t})$  data. Thirdly, the computations in WinBUGS, which is where the analysis was done, did not handle the binomial model. For these three reasons, I only use  $z_{i,t}$  equal an indicator for finding plague as plague data. The observation model is defined as:

$$z_{i,t} = \begin{cases} 1, & \text{if plague is observed for site } i \text{ at time } t. (y_{i,t} > 0) \\ 0, & \text{if plague is not observed for site } i \text{ at time } t. (y_{i,t} = 0, n_{i,t} > 0) \\ \text{undefined} & \text{if no plague data are available for site } i \text{ at time } t. (n_{i,t} = 0) \end{cases}$$

It is now natural to use a Bernoulli distribution:

$$z_{i,t} \sim \text{Bernoulli}(p_{i,t})$$

where  $p_{i,t}$  is the probability for finding plague for site  $i$  at time  $t$  if one would try.

The occupancy data also has observation errors that should be taken into account. To handle this, I have made an observation model for occupancy as well. The data is now given as

$$\phi_{i,t}^{\text{observed}} = \begin{cases} \{0, 1, \dots, 30\}, & \text{if estimate for occupancy are made for site } i \text{ at time } t. \\ \text{undefined} & \text{if no occupancy data are available for site } i \text{ at time } t. \end{cases}$$

As mentioned in the data description, the assumption that the number of burrows tested for site  $j$  at time  $t$ ,  $m_{j,t}$ , equals 30, seem to be incorrect in some cases. In addition I have some times taken the mean of two, three or even four estimates for primary squares. In other words,  $m_{j,t}$  is likely to be more than 30. My observation model is meant to give a likely variance to the observation given the true occupancy. Having more samples, would decrease this variance. However, I also mentioned that there might be clustering of inhabited burrows which would quite possibly give a positive correlation between the  $m_{j,t}$  samples.

This would increase the uncertainty of the data. For my observation model, I choose to assume that I have  $m_{j,t} = 30$  uncorrelated samples, and my model is given as

$$\phi_{j,t}^{\text{observed}} \sim \text{Bin}(30, \phi_{j,t})$$

where  $\phi_{j,t}$  is the occupancy equal to the proportion of occupied burrows for site  $j$  at time  $t$ . As mentioned this might have bigger or lower observation error than the reality, but it gives an approximate observation error.

The biological processes for  $p_{i,t}$  and  $\phi_{i,t}$  are modeled on logit scale as AR(1) processes over time, with noise terms modeled as Gaussian conditional autoregressive (GCAR) models, describing the spatial structure:

$$\begin{aligned} p_{i,t} &= \text{logit}^{-1}(p_{i,t}^*) \iff p_{i,t}^* = \text{logit}(p_{i,t}) \\ p_{i,t}^* &= \mu_{i,t}^p + \gamma_p * (p_{i,t-1}^* - \mu_{i,t-1}^p) + E_{i,t} & t \geq 2 \\ \mathbf{E}_t &\overset{i.i.d}{\sim} \text{GCAR}(\sigma_E, \rho_E) & \begin{array}{l} \sigma_E \text{ is the standard deviation} \\ \rho_E \text{ is the correlation parameter} \end{array} \end{aligned}$$

with  $\mathbf{E}_t = (E_{1,t}, \dots, E_{N,t})'$

$$\begin{aligned} \phi_{i,t} &= \text{logit}^{-1}(\phi_{i,t}^*) \iff \phi_{i,t}^* = \text{logit}(\phi_{i,t}) \\ \phi_{i,t}^* &= \mu_t^\phi + \gamma_\phi(\phi_{i,t-1}^* - \mu_{t-1}^\phi) + U_{i,t} & t \geq 2 \\ \mathbf{U}_t &\overset{i.i.d}{\sim} \text{GCAR}(\sigma_U, \rho_U) & \begin{array}{l} \sigma_U \text{ is the standard deviation} \\ \rho_U \text{ is the correlation parameter} \end{array} \end{aligned}$$

with  $\mathbf{U}_t = (U_{1,t}, \dots, U_{N,t})'$

For the simplest model,  $p_{i,t}$  and  $\phi_{i,t}$  are independent with expectations:

$$\begin{aligned} \mu_t^p &\overset{iid}{\sim} \text{Normal}(0, \sigma_{\mu^p}^2), \quad \mu_{i,t}^p = \mu_t^p \\ \mu_t^\phi &\overset{iid}{\sim} \text{Normal}(0, \sigma_{\mu^\phi}^2) \end{aligned}$$

but for the other models,  $\mu_{i,t}^p$  is defined as:

$$\mu_{i,t}^p = \beta' \mathbf{X},$$

using different covariates  $\mathbf{X}$  for each model including climatic factors depending on time only and occupancy. These models are illustrated in figure 4.

I have attempted to use an AR(2) process to describe the process of plague, but this was too computational demanding for my computer. The AR(2) process is of interest for ecologists, since an AR(2) process often may imply that important interacting species have been ignored.

Adjustments were made for  $t = 1$  to achieve stationarity. I have shown how to do this for  $p_{i,t}$  in the appendix. Similar adjustments were done for  $\phi_{i,t}$

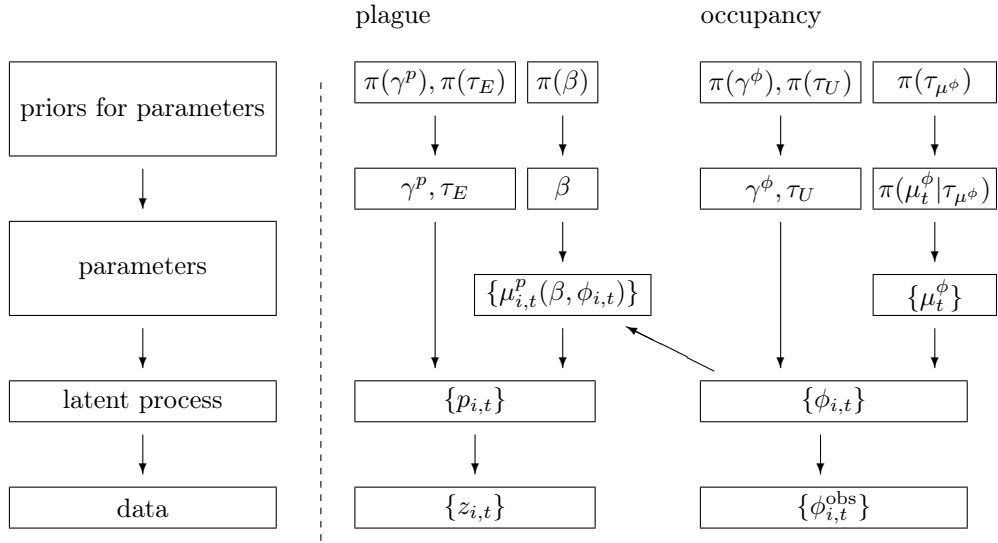


Figure 4: This figure is a sketch of the model. It shows priors, parameters, latent processes and observations to illustrate the Bayesian hierarchical model.  $\tau_E$  and  $\tau_U$  are precisions  $\frac{1}{\sigma_E^2}$  and  $\frac{1}{\sigma_U^2}$ .

**Definition: GCAR( $\sigma, \rho$ )**

Define GCAR( $\sigma$ ) to be the Gaussian Markov random field (Gaussian MRF), defined by

$$\mathbf{X} \sim \text{MVN}(0, \sigma^2 \mathbf{Q}^{-1}(\rho)), \quad \text{MVN} = \text{multivariate normal distribution}$$

with

$$\mathbf{Q}(\rho) = \rho \mathbf{P} + (1 - \rho) \mathbf{I}$$

$$P_{i,j} = \begin{cases} m_i, & \text{if } i = j \\ 1, & \text{if } i \sim j \\ 0, & \text{otherwise.} \end{cases}$$

## Inference

By using the Bayesian paradigm, I am interested in the posterior densities for each parameter of interest. These posterior densities will be estimated through simulations since the structure is complex. I will also use an information criterion which could help evaluate models and possibly to do model selection. These two combined will give a basis for my inference.

## MCMC

Markov chain Monte Carlo (MCMC) is a general method for sequentially sampling random quantities from an approximate distribution. These quantities are updated sequentially conditioned on each other, and thus corrected for each draw. In this way the approximate distribution converges to the target distribution, which is what one wants to sample from. When one assumes to have reached convergence, samples can be assumed to come from the target distribution.

To reach the target distribution, one uses a burn-in. This is a sample size that is drawn before one starts collecting samples as an approximation of the target distribution.

Let  $\theta$  bet the collection of all random quantities, and let  $\theta_i$  represent one or some. Let the burn-in be  $B$  and the sample size be  $M - B$ , then the MCMC method can be expressed as follows:

```
i)   initial values are set in  $\theta^{(0)}$ .
ii)  Letting the samples converge
      for(m in 1:B)
        for( j in 1:J )
          draw  $\theta_j^{(m)}$  given  $\theta_{1:(j-1)}^{(m)}, \theta_{(j+1):J}^{(m-1)}$  and the data
iii) Collecting/Saving samples
      for(m in (B+1):M)
        for( j in 1:J )
          draw  $\theta_j^{(m)}$  given  $\theta_{1:(j-1)}^{(m)}, \theta_{(j+1):J}^{(m-1)}$  and the data
```

Also there is sometimes used a thinning. This means to skip a certain number of samples to avoid problems with storing, but also to decrease correlations between samples.

A particular MCMC algorithm is the Gibbs sampler. This algorithm uses the conditional distribution for updating the  $\theta$ 's.

I have used a burn-in of 1 000 and my sample size is 9 000. It took days to run these 10 000 iterations. I have shown convergence and stability of the parameters of interest in the appendix.

For more details on MCMC, see e.g. Robert and Casella (2004).

## Model evaluation by DIC

I would like to do model selections to decide which factors I should report as important or significant. In other settings with simpler models and no missing data, AIC is often used. However, in complicated situations, such as this analysis, AIC cannot be used.

The usual model criterion used in many frequentistic analyzes, AIC, is defined as

$$\text{AIC} = -2\log\text{likelihood}(\theta) + 2p, \quad p = |\theta|$$

and is minimized wrt models. AIC and other information criteria usually depends on calculating the log likelihood for a given model. However, in my analysis the log likelihoods depend on random parameters and are complicated to compute. For such situations, there is another information criterion, DIC, calculated by BUGS without much difficulty. DIC is defined very similarly to the AIC, but is designed for Bayesian hierarchical models:

$$\text{DIC} = -2\log\text{likelihood}(\hat{\theta}) + 2pD, \quad \begin{array}{l} pD = \text{effective number of parameters} \\ \hat{\theta} = \text{mean of the parameters } \theta. \end{array}$$

The calculations of DIC is done by first calculating

$$\begin{aligned} \bar{D} &= \overline{-2\log \text{likelihood}(\theta)} \\ \hat{D} &= -2\log \text{likelihood}(\bar{\theta}) \end{aligned}$$

Then one calculates

$$\begin{aligned} \text{pD} &= \text{'the effective number of parameters'} = \bar{D} - \hat{D} \\ \text{DIC} &= \bar{D} + \text{pD} = \hat{D} + 2\text{pD} \end{aligned}$$

DIC is intended to be a generalization of AIC. The definitions of AIC and DIC are very similar, but instead of using the punishment  $2p$ , DIC use  $2\text{pD}$ . If DIC was to be used on a non-hierarchical model, it would give similar results as AIC. However, with hierarchical models as I use, the information criteria will be different from each other.

Unfortunately, one cannot simply trust DIC statistics to choose a winner model. I found that for different seeds the DIC varied more than I find pleasant. To give stable results, the DIC need more iterations than the time available permitted.

One should notice that DIC is sensitive to how you write your model. To illustrate the WinBUGS manual gives the following example:

```
tau ~ dgamma(0.001, 0.001)
```

and

```
log.tau ~ dunif(-10, 10); log(tau) <- log.tau
```

The two definition of  $\tau$  (`tau`) is theoretically equivalent, but in BUGS they will appear different. Calculating  $\hat{D}$  will give different result. In the first example  $\bar{\tau}$  will be used, but in the second example  $\overline{\log \tau}$  will be used.

As expected, DIC is also sensitive to priors used. This is visualized in the appendix for model 0a.

Considering these weaknesses, I would prefer to use another information criteria. E.g. I could have used the number of correct estimates  $\hat{z} = I(p_{i,t} > 0.5)$  of  $z$  for cross validation. The model returning the highest count, would be the preferred model.

## 4 BUGS

Bayesian inference Using Gibbs Sampling (BUGS) is a simulation tool that does MCMC with a simple interface. If it finds conditional distributions, it does Gibbs sampling. The model used is easy to control, in the sense that it is easy to read the model code used by BUGS. BUGS also allows the user to change the model quick and easy by simply changing priors, conditional distributions or a parameter value. However, it is not computationally fast, and error messages are often difficult to understand. When I began this project, the choice between BUGS and other programming languages, such as R or C++, was a choice of using much time on the programming or using much time on the simulations. As we were uncertain of how many models we would choose, the choice was to use BUGS for simplicity.

There are two versions of BUGS, namely WinBUGS<sup>4</sup> and OpenBUGS<sup>5</sup>. These are often considered to be equal, and there are not very many differences. However, there are some. The development are now mainly focused on OpenBUGS. Since it often is claimed that these two versions of BUGS are essentially equal, one might think they behave equally. This was not true for my models. On the OpenBUGS home page, one finds a list of differences. The first thing listed is the updating routines.

Figure 5 shows that the OpenBUGS simulations gave few updates and no convergence for both  $\beta$  and  $\gamma_p$ , and the standard deviation  $\sigma_E$  increased exponentially. The WinBUGS simulations seemed to converge and gave many updates for both  $\beta$  and  $\gamma_p$ , and  $\sigma_E$  stayed in the region [2, 3].

Looking at these results, the choice of which BUGS version to use were not difficult. WinBUGS were the natural choice since convergence and stability was much better than that of OpenBUGS - at least in my case. It should be mentioned that there are added possibilities for changing the updating routines in OpenBUGS. I used the default settings, but it might be possible to get good results for OpenBUGS if one changes the updating routines. A easy way out was to use the default setting of WinBUGS instead of each time changing the settings for updating routines in OpenBUGS.

For an introduction to BUGS, see the WinBUGS manual, the OpenBUGS manual or Doing Bayesian Data Analysis: A tutorial with R and BUGS, John K. Kruschke, 2011. Kruschke and John (2010)

The BRugs package for calling BUGS from R, used throughout the book by Kruschke, have been removed from the CRAN repository. One could get old versions of BRugs from the cran archives, but there is another package, R2WinBUGS, that seem to work even better.

---

<sup>4</sup>Lunn et al. (2007)

<sup>5</sup>Lunn et al. (2011)

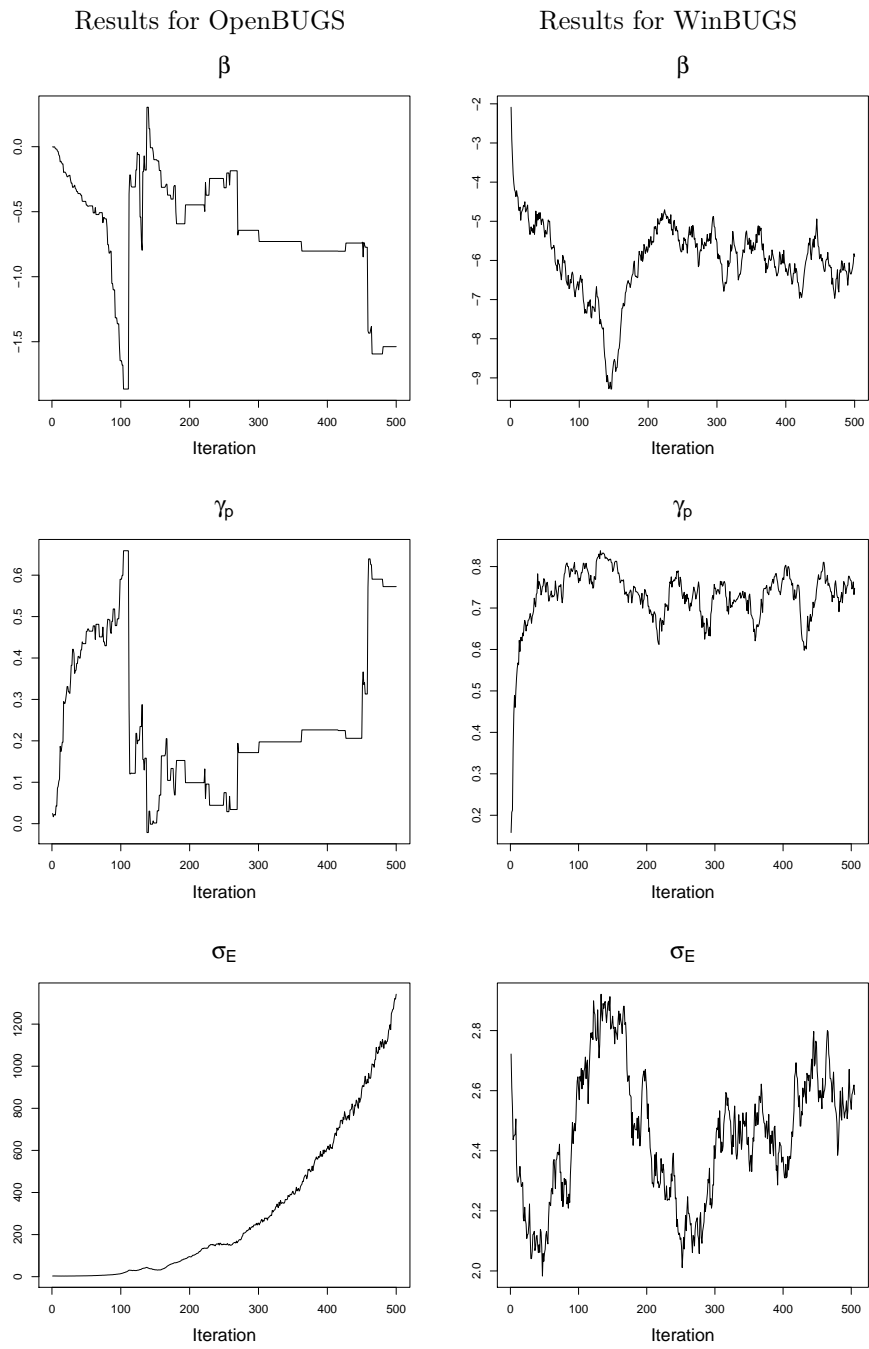


Figure 5: This figure shows the first 500 iterations for model 0b in both OpenBUGS and WinBUGS. The model is the model with only intercept defining the expectation for  $p_{i,t}^*$ ,  $\mu_t^p = \beta$

## GeoBUGS

GeoBUGS is a package in BUGS, which is especially designed for spatial models and visualizations on maps. It grants the opportunity for simulating spatial models such as various multinormal models and a double exponential model.

I have used the function `car.normal()` to simulate the GCAR noise, which is defined in the method section. It demands the precision,  $\tau = 1/\sigma^2$ , and the three vectors defining the spatial structure, (i) number of neighbors for every site, (ii) adjacent areas to every site and (iii) unnormalized weights. All weights equals 1 in my case. The `car.normal()` function simulates  $\text{GCAR}(\rho, \tau)$  processes with  $\tau$  specified by the user. However,  $\rho$  is not shown to the user.

## Trap

For flat priors I got an error with a TRAP message and it was not possible to proceed the simulations. The WinBUGS user manual gives several advices to handle this. The only reasonable solution of these in my case seemed to be more informative priors. I used strict priors on  $\tau_E = \frac{1}{\sigma^2}$  and  $\beta$ . The priors for  $\beta$  is unbiased, except for the intercept,  $\beta_1$ , in a few cases.

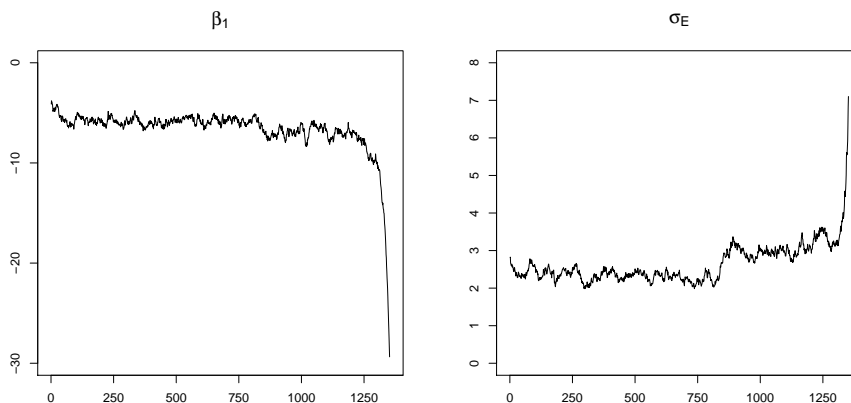


Figure 6: The figure shows an example of a simulation that ended in a TRAP message

Figure 6 shows an example of a simulation that ended with a TRAP message. This example is the model that needed even stricter priors than the other models, but with the prior used by other models. The figure shows somewhat stable simulations for both  $\beta_1$  and  $\sigma_E$  before they diverged at the end. To get simulations to proceed until 10 000 iterations, I used a stricter prior for  $\beta_1$  centered on -5 with a variance of 2. This prior seemed reasonable, when looking at results from other simulations.



## Achieving DIC

The accessibility is the main reason for me to report DIC values. It is obtained as follows:

1. Run a Burn-in
2. Press set in the DIC Tool window
3. Run the sample size you want to use
4. Press DIC in the DIC Tool Window. This opens the DIC window.

The DIC window gives a table with the statistics  $\bar{D}$ ,  $\hat{D}$ , pD and DIC, defined in Model evaluation by DIC, for every separately defined variable (in my case `z` (referring to plague) and `occ.obs` (referring to occupancy)). The statistics is also given for the `total` which is sums of the statistics above.

## 5 Results

### The models

The occupancy model is combined with several models for plague. The differences between these plague models is the definition of covariates,  $\mathbf{X}$ . As mentioned occupancy ( $\phi_{i,t}$ ) and climatic covariates is expected by biologists to be the most important factors to explain presence and prevalence of plague. Kyrre Linné Kausrud et al<sup>6</sup> suggest April to be the most important for climatic influence on the great gerbils. I therefore test both average temperature and average humidity in April. I also test the average temperature for the coldest and the warmest month, January and July respectively.

### Covariates

$\phi_{i,t}^*$ : logit of the occupancy at site  $i$  and year  $t$ .

$\mathbf{T}_{1,t}$ : (standardized) average temperature for January, in year  $t$ .

$\mathbf{T}_{4,t}$ : (standardized) average temperature for April, in year  $t$ .

$\mathbf{T}_{7,t}$ : (standardized) average temperature for July, in year  $t$ .

$\mathbf{H}_{4,t}$ : (standardized) average humidity for April, in year  $t$ .

Model	Occupancy	Temperature			Humidity
	$\phi_{i,t}^*$	$\mathbf{T}_{1,t}$	$\mathbf{T}_{4,t}$	$\mathbf{T}_{7,t}$	$\mathbf{T}_{4,t}$
0a					
0b					
1	★				
2	★		★		
3	★	★	★		
4	★	★	★	★	
5	★	★	★		★

Table 1: This table shows which covariates I used for the models. If a covariate is included in a model, this is shown with the symbol ★.

0a: independent expectations for every year

$$\mu_{i,t}^p = \mu_t^p$$

0b: equal expectation for every year

$$\mu_{i,t}^p = \mu^p$$

### Reporting results

The box plots for  $\mu_t^p$  and  $\mu_t^\phi$  are made in  $\mathbf{R}^7$  with the `boxplot` function. The estimated densities for parameters are made in  $\mathbf{R}$  with the function `density`.

In the appendix section "Trace plots" I give trace plots that shows convergence. All parameters shown seem to have converged, but I would have liked more samples of the intercept,  $\beta_1$ .

<sup>6</sup>Linné Kausrud et al. (2007)

<sup>7</sup>R Development Core Team (2010)

## Occupancy

The results for the occupancy processes were approximately the same for each model. As shown in table 2, DIC varied very little between models. On the graphical representation of the results, there were very difficult to spot any differences. To illustrate the occupancy results I have chosen the Model 0b, in which occupancy and plague is modeled independently.

We see from figure 7 distinct differences of  $\mu_t^\phi$ . This might imply that the

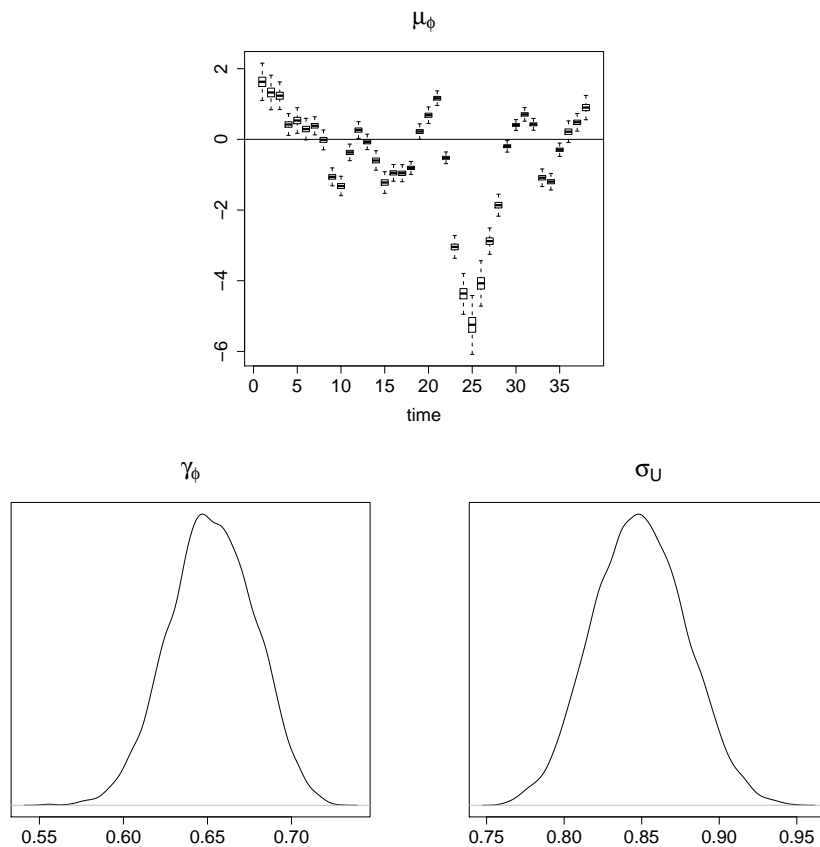


Figure 7: Results for occupancy.

simulated occupancy would give good estimations for the missing data. And in turn that  $\beta_2$  in the plague model might be well estimated.

## DIC

Table 2 shows the DIC values obtained for all models. Unfortunately, the DIC values were too unstable to only rely on them. Model 3 was ran a second time with a different seed. This returned a total DIC of 8713.690, which is unpleasantly different from the DIC in the first attempt which is shown in table 2. Keeping this in mind, we see that DIC does not choose a model very strongly. This illustrates that DIC in some cases should not be trusted. It may be improved by increasing the sample size. I was not able to do this in this analysis due to limitations of the computational force. To run one model, I used more than 2 days. To increase the number of iterations to say 100 000, would probably take more than 20 days for each simulation. This would not be possible.

Model	Plague		Occupancy		Total	
	pD	DIC	pD	DIC	pD	DIC
0a <sup>1</sup>	<i>142.747</i>	<i>516.054</i>	<i>962.308</i>	<i>8096.390</i>	<i>1105.050</i>	<i>8612.450</i>
0b <sup>1</sup>	<i>138.623</i>	<i>624.355</i>	<i>958.379</i>	<i>8092.350</i>	<i>1097.000</i>	<i>8716.710</i>
1	133.485	650.483	956.678	8096.430	1090.160	8746.910
2	142.604	614.731	955.993	8095.200	1098.600	8709.930
3	140.323	625.759	957.140	8096.310	1097.460	8722.060
4	170.289	<b>591.948</b>	956.771	8094.660	1127.060	<b>8686.610</b>
5 <sup>1</sup>	<i>165.174</i>	<i>593.427</i>	<i>957.709</i>	<i>8095.960</i>	<i>1122.880</i>	<i>8689.390</i>

<sup>1</sup>Model 0a, 0b and model 5 are not comparable to the other models. These models were not possible to run under the restrictions used for models 1, 2, 3 and 4. This has to do with the model restrictions and priors used.

Table 2: This table shows DIC results.

Though DIC does give a clear winner model, model 4 seem to be chosen. However, the uncertainty of DIC results demands a closer investigation of which model should be used.

When I ran model 5 with the same limitations and priors as the other models 1,2,3 and 4, simulations ended in a TRAP message before enough iterations were ran. To avoid this, I had to use a stricter prior for the intercept,  $\beta_1$ , but this might mean that the DIC are not comparable with the other models. Model 5 is therefor investigated further and compared with the model with the best DIC scores, model 4.

Model 0a and model 0b are models of little interest as a final model. They do not give information to the biologists on the process of sylvatic plague. Also, I had to use less restrictions on these models. These models ended in trap messages when I demanded stationarity. However, simulations showed stationarity to be very likely.

## Model 0a and Model 0b

The interest of model 0a and model 0b lies in (1) the estimated level of plague risk for each year and (2) the fit  $\hat{D}$ , defined in "Model evaluation by DIC", in the sense of how much could possibly be explained by yearly factors. In this way, these models are in some sense reference models.

- Model 0a has independent  $\mu_t^p$  and will give a best possible fit excluding occupancy.
- Model 0b is the opposite. This will be the least explained model, but provides information on how much is explained simply by the intercept.

Combined these two models gives some information on the power. If the fit ( $\hat{D}$ ), is very different for the two models, there might be much that could be explained. If they are practically equal, there is nothing to explain by covariates on  $\mu_t^p$ . These models are independent from occupancy, and the interesting fit is the plague fit given in table 3.

For model 0a, the DIC results varied much for all reasonable priors<sup>8</sup>. The result given for model 0a in table 3 is the best fit of all priors tested.

Model	$\hat{D}$
0a	230.560
0b	347.110

Table 3: This table shows  $\hat{D}$  for plague.

The figure 8 and the figure 9 shows results for model 0a and for model 0b, respectively. One can see that both models gives high values for  $\gamma_p$ . The standard deviation,  $\sigma_E$ , is estimated to quite similar the regions. Also, one sees the expectation of  $p_{i,t}^*$  to be estimated below zero.  $p_{i,t}^*$  below zero is equivalent with  $p_{i,t}$  below 0.5.

Figure 8 shows some structure of  $\mu_{i,t}^p$  for model 0a. Since the measure of fit,  $\hat{D}$ , for model 0a is quite small compared to that of model 0b, one could possibly be able to explain some of this. One might say that the power is high enough.

---

<sup>8</sup>See details in the Appendix section "Model 0a"

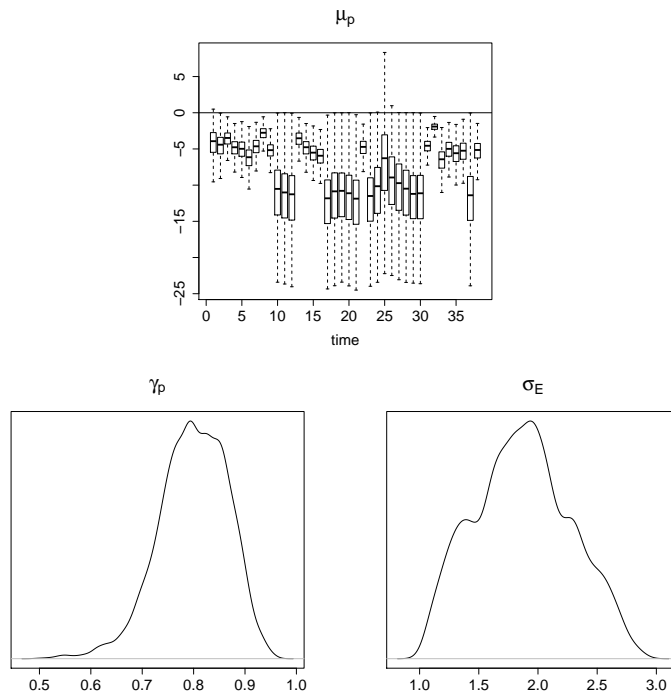


Figure 8: Results for model 0a

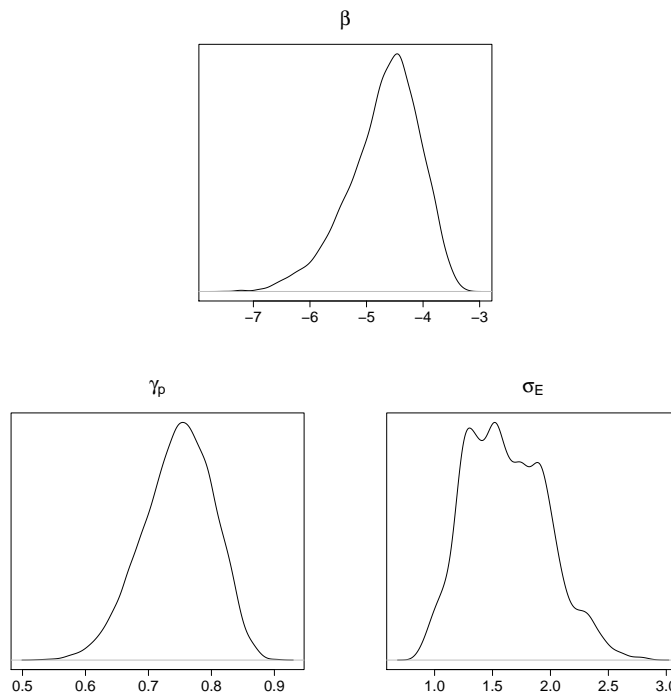


Figure 9: Results for model 0b

## Model 5

For model 5 the expectation for  $p_{i,t}^*$  is defined as

$$\mu_{i,t}^p = \beta_1 + \beta_2 \phi_{i,t}^* + \beta_3 \mathbf{T}_{1,t} + \beta_4 \mathbf{T}_{4,t} + \beta_5 \mathbf{H}_{4,t}$$

To evaluate model 5, I compare the model with the model returning the best DIC values in table 2, model 4. Table 4 shows that model 4 gives better (smaller) DIC values than model 5 with equal priors as well.

Figure 10 shows the estimated posterior density for  $\beta_5$  for model 5. This is the parameter that gives the effect of humidity. It is centered close to zero with many samples drawn from both the positive region and the negative region. This gives a valid reason to reject the model or at least to remain skeptic to including humidity.

Model	Plague		Occupancy		Total	
	pD	DIC	pD	DIC	pD	DIC
4	165.033	<b>586.073</b>	959.858	8096.680	1124.890	<b>8682.750</b>
5	165.174	593.427	957.709	8095.960	1122.880	8689.390

Table 4: This figure shows DIC values for model 5 and model 2 with equal strict priors.

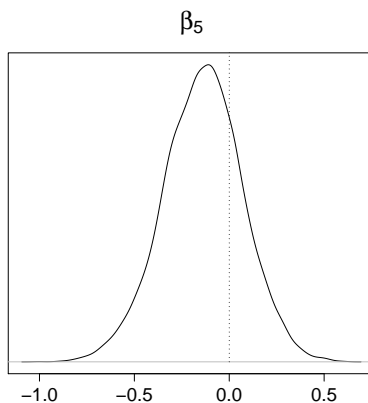


Figure 10: We see from this that it is not unreasonable with  $\beta_5 = 0$ .

## Model 4 - the best model according to the DIC

Model 4 is an expansion of model 1, model 2 and model 3. The expectation for  $p_{i,t}^*$  is defined as

$$\mu_{i,t}^p = \beta_1 + \beta_2 \phi_{i,t}^* + \beta_3 \mathbf{T}_{1,t} + \beta_4 \mathbf{T}_{4,t} + \beta_5 \mathbf{T}_{7,t}$$

As mentioned, DIC results behaved somewhat unstable for my 9 000 iterations. Since estimated marginal posterior densities for the parameters are more stable, one should look at these when considering which model is the best. The estimated marginal posterior densities gave similar results for all explanatory variables for all models.

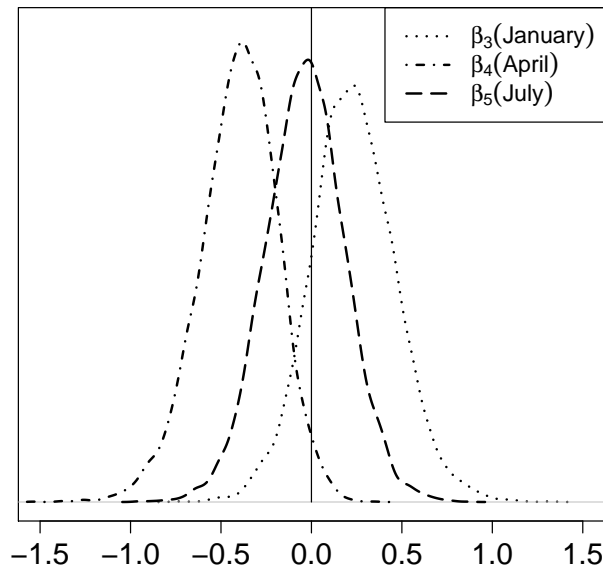


Figure 11: Temperature

The difference between model 3, model 4 and model 5 is only related to  $\beta_5$ . For model 3,  $\beta_5$  is excluded. For model 4 and model 5,  $\beta_5$  is related to average temperature in July and average humidity in April, respectively. In "Model 5", figure 10 illustrated that the average humidity in April had little or no influence on plague. As for the average temperature of July, figure 11 shows that  $\beta_5$  has the mode very close to zero. These estimated densities strongly implies model 3 to be better than both of the models 4 and 5.

For  $\beta_3$ , approximately 15 % of the simulated samples are less than zero. This indicates that the effects of average temperature in January is quite uncertain as well. The effects of average temperature in April is much more significant with less than 2.5 % of simulated  $\beta_4$  above zero, which means that the 95 % credibility interval not includes zero. Now model 2, which only includes average temperature of April as climatic influence, seem to be the best model.



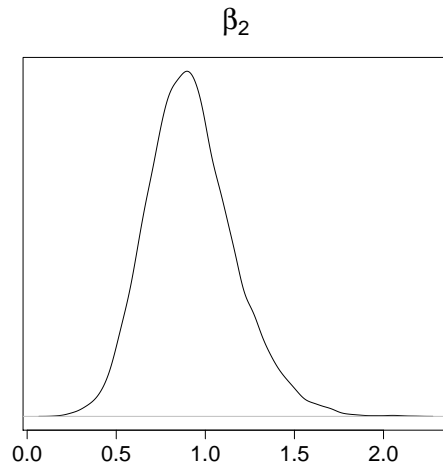


Figure 12: Occupancy

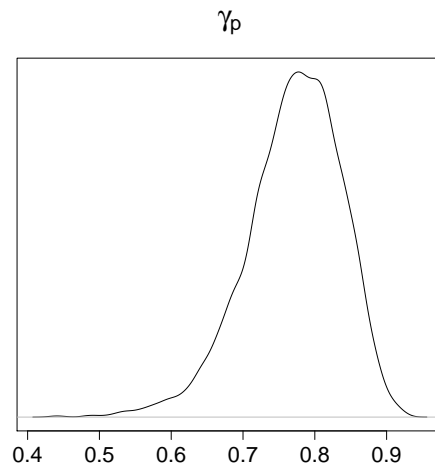


Figure 13: The AR parameter  $\gamma_p$  which gives the temporal structure.

Even more evident than the effect of average temperature of April, is that of occupancy illustrated by figure 12. If there is a high occupancy, there is more likely to be plague present among the tested rodents. This indicates that the process of plague is closely related to the occupancy process.

Figure 13 shows that the AR(1) parameter  $\gamma_p$  is quite big. This may indicate that if there is a higher prevalence of plague than expected at site  $i$  a given year, there is more likely to be found plague the next year as well.

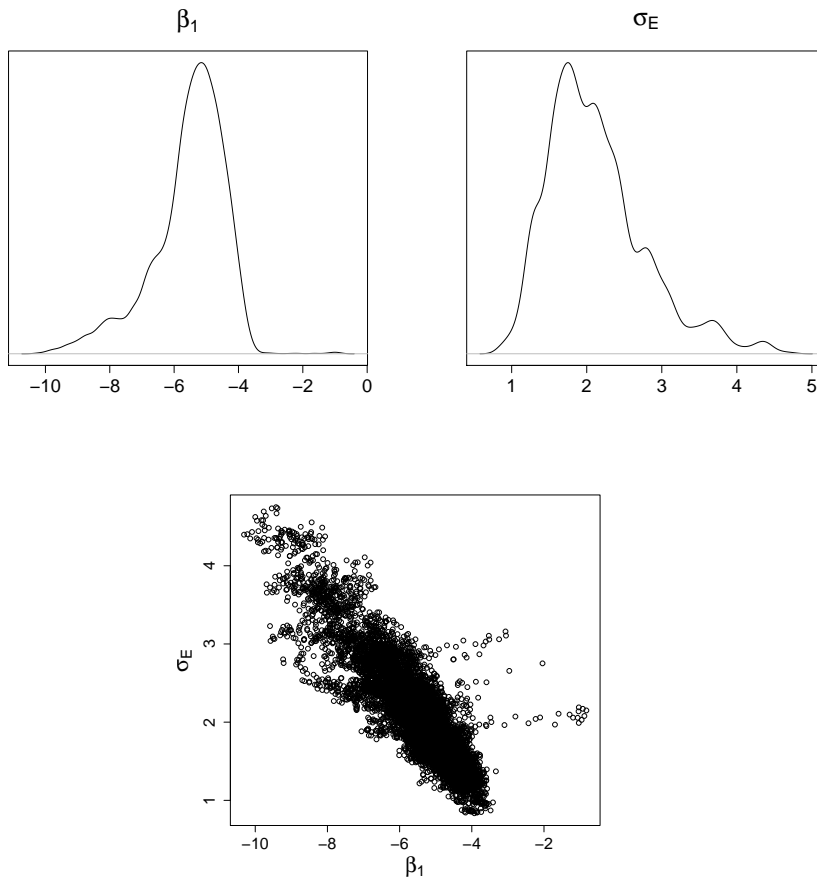


Figure 14: These graphics shows results for the intercept and the standard deviation for  $p_{i,t}^*$ .

### $\beta_1$ and $\sigma_E$

Figure 14 shows results for  $\beta_1$ , the intercept for the process of  $p_{i,t}^*$ , and  $\sigma_E$ , the standard deviation for GCAR in plague process. It seems  $\beta_1$  and  $\sigma_E$  is strongly correlated. This may be why these may appear less smooth in their estimated posterior densities.

## 6 Summary / Discussion

In the desert Saryyesik-Atyrau, Kazakhstan, sylvatic plague has been closely monitored for many years to minimize the risk of infections among humans. A huge field was divided into squares. Every year some squares were tested. Usually, these squares were chosen randomly, but if there was a report of sick animals in a certain square, this square was tested. This resulted in missing data not being missing at random. Ignoring this, may give bigger estimates for plague risk than the reality. I do not know which or even how many squares were randomly chosen, but luckily it is indicated that there was few instances of such reports, and it is usually ignored. Based on the indications that there was not many cases of such reports, I have also chosen to ignore this.

Through a Bayesian hierarchical model, I have analyzed the process of sylvatic plague with respect to occupancy and climate. I have also included a spatiotemporal structure. By using such a model, I handled the general uncertainty introduced by missing data quite well. This uncertainty is also often ignored.

The DIC results, which were quite unstable, prefers the model with average temperature from all three months tested, January, April and July, but investigating the estimated marginal posterior densities, I found both January and July to be insignificant. The fact that the mean temperature for April seems to be more important than both January and July, may have several reasons. Two of which is that (i) April, or possibly the spring in general, may be the most important period for biological reasons and that (ii) it may be more important to have a long period of summer-like conditions. Stenseth et al. (2006) argue that (ii) may be important for flea activity and thus important for the risk of plague infection on great gerbils. By the arguments in Stenseth et al. (2006), one should expect a positive estimate for this  $\beta$ -parameter. However, for my model the average temperature of April is estimated to have a negative correlation with plague. To test (ii) one could have used the number of days with temperature above a certain threshold as an explanatory variable as well.

As for humidity, both DIC and the estimated posterior marginal posterior densities indicates that average humidity of April has no significant influence on plague.

The analysis clearly indicates that occupancy has influence on the process of sylvatic plague. If there is a high occupancy, it seems the probability for finding plague among the trapped rodents is increased.

Also, an infected area is more likely than an uninfected area to be infected the next year.

MCMC computations have been done in WinBUGS. Unfortunately, WinBUGS had several limitations on the computations. Most of these could probably be improved by doing the computations in R or C++. Some limitations was related to the updating routines. E.g.: I attempted to expand the process of sylvatic plague from an AR(1) model to an AR(2) model. BUGS did not handle this expansion. Other limitations were of visualization. I could not investigate spatial correlation.

Another expansion that would be interesting is the expansion of including flea data. This is assumed by biologists to be of great importance, but the uncertainty of how such data were collected and the computationally demanding process prevented me from including this in my analysis.

# Appendix

## A BUGS models

This section gives models 1, 2, 3, 4 and 5, as defined in WinBUGS.

The models are easy to read as it is written straight forward.

Priors are given by e.g.

```
 $\tau_E \sim \text{Gamma}(2, 0.2)$ :  
tau.E ~ dgamma(2,0.2)
```

or

```
 $\beta_i \sim \text{N}(0, 10)$ ,  $i = 1, \dots, 5$ :  
for(j in 1:2) {  
  beta.p[j] ~ dnorm(0, 0.1)  
}
```

The observation model for plague is given by

```
 $z_{i,t} \sim \text{Bernoulli}(p_{i,t})$ ,  $i \in \mathcal{D}_t^p$ :  
for(t in 1:T) {  
  for(i in 1:N.z[t]) {  
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])  
  }  
}
```

## A.1 Model 1

```

model{
for(t in 1:T) {
  for(i in 1:N.z[t]) {
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])
  }
}

for (i in 1:N) {
  logit(p[i,1]) <- x.p[i,1]
  x.p[i,1] <- mu.p[i,1] + c.p*E[ i ]
  mu.p[i,1] <- beta.p[1] + beta.p[2]*x.occ[i,1]
}
mean.mu.p[1] <- mean(mu.p[1:N,1])
for(t in 2:T) {
  for (i in 1:N) {
    logit(p[i,t]) <- x.p[i,t]
    x.p[i,t] <- mu.p[i,t] + gamma.p*(x.p[i,t-1] - mu.p[i,t-1]) + E[i + N*(t-1)]
    mu.p[i,t] <- beta.p[1] + beta.p[2]*x.occ[i,t]
  }
  mean.mu.p[t] <- mean(mu.p[1:N,t])
}

for(j in 1:2) {
  beta.p[j] ~ dnorm(0, 0.1)
}

gamma.p ~ dunif(-1,1)
c.p <- 1/sqrt(1 - pow(gamma.p , 2))

for(t in 1:T) {
  E[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.E)
}
tau.E ~ dgamma(2,0.2)
sigma.E <- sqrt(1/tau.E)

for(t in 1:T) {
  for(i in 1:N.occ[t]) {
    occ.obs[ind.occ[i,t],t] ~ dbin(occ[ind.occ[i,t],t], 30)
  }
}

for (i in 1:N) {
  logit(occ[i,1]) <- x.occ[i,1]
  x.occ[i,1] <- mu.occ[1] + c.occ*U[ i ]
}
for(t in 2:T) {
  for (i in 1:N) {
    logit(occ[i,t]) <- x.occ[i,t]
    x.occ[i,t] <- mu.occ[t] + gamma.occ*(x.occ[i,t-1] - mu.occ[t-1]) + U[i + N*(t-1)]
  }
}

for(t in 1:T) {
  mu.occ[t] ~ dnorm(0,tau.mu.occ)
}
tau.mu.occ ~ dgamma(0.1,0.02)

gamma.occ ~ dunif(-1,1)
c.occ <- 1/sqrt(1 - pow(gamma.occ , 2))

for(t in 1:T) {
  U[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.U)
}
tau.U ~ dgamma(2,0.2)
sigma.U <- sqrt(1/tau.U)
}

```

## A.2 Model 2

```

model{
for(t in 1:T) {
  for(i in 1:N.z[t]) {
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])
  }
}

for (i in 1:N) {
  logit(p[i,1]) <- x.p[i,1]
  x.p[i,1] <- mu.p[i,1] + p.start*E[ i ]
  mu.p[i,1] <- beta.p[1] + beta.p[2]*x.occ[i,1] + beta.p[3]*temp4[1]
}
mean.mu.p[1] <- mean(mu.p[1:N,1])
for(t in 2:T) {
  for (i in 1:N) {
    logit(p[i,t]) <- x.p[i,t]
    x.p[i,t] <- mu.p[i,t] + gamma.p*(x.p[i,t-1] - mu.p[i,t-1]) + E[i + N*(t-1)]
    mu.p[i,t] <- beta.p[1] + beta.p[2]*x.occ[i,t] + beta.p[3]*temp4[t]
  }
  mean.mu.p[t] <- mean(mu.p[1:N,t])
}

for(j in 1:3) {
  beta.p[j] ~ dnorm(0, 0.1)
}

gamma.p ~ dunif(-1,1)
p.start <- 1/sqrt(1 - pow(gamma.p , 2))

for(t in 1:T) {
  E[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.E)
}
tau.E ~ dgamma(2,0.2)
sigma.E <- sqrt(1/tau.E)

for(t in 1:T) {
  for(i in 1:N.occ[t]) {
    occ.obs[ind.occ[i,t],t] ~ dbin(occ[ind.occ[i,t],t], 30)
  }
}

for (i in 1:N) {
  logit(occ[i,1]) <- x.occ[i,1]
  x.occ[i,1] <- mu.occ[1] + occ.start*U[ i ]
}
for(t in 2:T) {
  for (i in 1:N) {
    logit(occ[i,t]) <- x.occ[i,t]
    x.occ[i,t] <- mu.occ[t] + gamma.occ*(x.occ[i,t-1] - mu.occ[t-1]) + U[i + N*(t-1)]
  }
}

for(t in 1:T) {
  mu.occ[t] ~ dnorm(0,tau.mu.occ)
}
tau.mu.occ ~ dgamma(0.1,0.02)

gamma.occ ~ dunif(-1,1)
occ.start <- 1/sqrt(1 - pow(gamma.occ , 2))

for(t in 1:T) {
  U[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.U)
}
tau.U ~ dgamma(2,0.2)
sigma.U <- sqrt(1/tau.U)
}

```

### A.3 Model 3

```

model{
for(t in 1:T) {
  for(i in 1:N.z[t]) {
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])
  }
}

for (i in 1:N) {
  logit(p[i,1]) <- x.p[i,1]
  x.p[i,1] <- mu.p[i,1] + c.p*E[ i ]
  mu.p[i,1] <- beta.p[1] + beta.p[2]*x.occ[i,1] + beta.p[3]*temp1[1] + beta.p[4]*temp4[1]
}
mean.mu.p[1] <- mean(mu.p[1:N,1])
for(t in 2:T) {
  for (i in 1:N) {
    logit(p[i,t]) <- x.p[i,t]
    x.p[i,t] <- mu.p[i,t] + gamma.p*(x.p[i,t-1] - mu.p[i,t-1]) + E[i + N*(t-1)]
    mu.p[i,t] <- beta.p[1] + beta.p[2]*x.occ[i,t] + beta.p[3]*temp1[t] + beta.p[4]*temp4[t]
  }
  mean.mu.p[t] <- mean(mu.p[1:N,t])
}

for(j in 1:4) {
  beta.p[j] ~ dnorm(0, 0.1)
}

gamma.p ~ dunif(-1,1)
c.p <- 1/sqrt(1 - pow(gamma.p , 2))

for(t in 1:T) {
  E[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.E)
}
tau.E ~ dgamma(2,0.2)
sigma.E <- sqrt(1/tau.E)

for(t in 1:T) {
  for(i in 1:N.occ[t]) {
    occ.obs[ind.occ[i,t],t] ~ dbin(occ[ind.occ[i,t],t], 30)
  }
}

for (i in 1:N) {
  logit(occ[i,1]) <- x.occ[i,1]
  x.occ[i,1] <- mu.occ[1] + c.occ*U[ i ]
}
for(t in 2:T) {
  for (i in 1:N) {
    logit(occ[i,t]) <- x.occ[i,t]
    x.occ[i,t] <- mu.occ[t] + gamma.occ*(x.occ[i,t-1] -mu.occ[t-1]) + U[i + N*(t-1)]
  }
}

for(t in 1:T) {
  mu.occ[t] ~ dnorm(0,tau.mu.occ)
}
tau.mu.occ ~ dgamma(0.1,0.02)

gamma.occ ~ dunif(-1,1)
c.occ <- 1/sqrt(1 - pow(gamma.occ , 2))

for(t in 1:T) {
  U[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.U)
}
tau.U ~ dgamma(2,0.2)
sigma.U <- sqrt(1/tau.U)
}

```



## A.4 Model 4

```

model{
for(t in 1:T) {
  for(i in 1:N.z[t]) {
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])
  }
}

for (i in 1:N) {
  logit(p[i,1]) <- x.p[i,1]
  x.p[i,1] <- mu.p[i,1] + c.p*E[ i ]
  mu.p[i,1] <- beta.p[1] + beta.p[2]*x.occ[i,1] + beta.p[3]*temp1[1] + beta.p[4]*temp4[1] + beta.p[5]*temp7[1]
}
mean.mu.p[1] <- mean(mu.p[1:N,1])
for(t in 2:T) {
  for (i in 1:N) {
    logit(p[i,t]) <- x.p[i,t]
    x.p[i,t] <- mu.p[i,t] + gamma.p*(x.p[i,t-1] - mu.p[i,t-1]) + E[i + N*(t-1)]
    mu.p[i,t] <- beta.p[1] + beta.p[2]*x.occ[i,t] + beta.p[3]*temp1[t] + beta.p[4]*temp4[t] + beta.p[5]*temp7[t]
  }
  mean.mu.p[t] <- mean(mu.p[1:N,t])
}

for(j in 1:5) {
  beta.p[j] ~ dnorm(0, 0.1)
}

gamma.p ~ dunif(-1,1)
c.p <- 1/sqrt(1 - pow(gamma.p , 2))

for(t in 1:T) {
  E[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.E)
}
tau.E ~ dgamma(2,0.2)
sigma.E <- sqrt(1/tau.E)

for(t in 1:T) {
  for(i in 1:N.occ[t]) {
    occ.obs[ind.occ[i,t],t] ~ dbin(occ[ind.occ[i,t],t], 30)
  }
}

for (i in 1:N) {
  logit(occ[i,1]) <- x.occ[i,1]
  x.occ[i,1] <- mu.occ[1] + c.occ*U[ i ]
}
for(t in 2:T) {
  for (i in 1:N) {
    logit(occ[i,t]) <- x.occ[i,t]
    x.occ[i,t] <- mu.occ[t] + gamma.occ*(x.occ[i,t-1] - mu.occ[t-1]) + U[i + N*(t-1)]
  }
}

for(t in 1:T) {
  mu.occ[t] ~ dnorm(0,tau.mu.occ)
}
tau.mu.occ ~ dgamma(0.1,0.02)

gamma.occ ~ dunif(-1,1)
c.occ <- 1/sqrt(1 - pow(gamma.occ , 2))

for(t in 1:T) {
  U[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.U)
}
tau.U ~ dgamma(2,0.2)
sigma.U <- sqrt(1/tau.U)
}

```

## A.5 Model 5

```

model{
for(t in 1:T) {
  for(i in 1:N.z[t]) {
    z[ind.z[i,t],t] ~ dbern(p[ind.z[i,t],t])
  }
}

for (i in 1:N) {
  logit(p[i,1]) <- x.p[i,1]
  x.p[i,1] <- mu.p[i,1] + c.p*E[ i ]
  mu.p[i,1] <- beta.p[1] + beta.p[2]*x.occ[i,1] + beta.p[3]*temp1[1] + beta.p[4]*temp4[1] + beta.p[5]*humid[1]
}
mean.mu.p[1] <- mean(mu.p[1:N,1])
for(t in 2:T) {
  for (i in 1:N) {
    logit(p[i,t]) <- x.p[i,t]
    x.p[i,t] <- mu.p[i,t] + gamma.p*(x.p[i,t-1] - mu.p[i,t-1]) + E[i + N*(t-1)]
    mu.p[i,t] <- beta.p[1] + beta.p[2]*x.occ[i,t] + beta.p[3]*temp1[t] + beta.p[4]*temp4[t] + beta.p[5]*humid[t]
  }
  mean.mu.p[t] <- mean(mu.p[1:N,t])
}

beta.p[1] ~ dnorm(-5,0.5)
for(j in 2:5) {
  beta.p[j] ~ dnorm(0, 0.1)
}

gamma.p ~ dunif(-1,1)
c.p <- 1/sqrt(1 - pow(gamma.p , 2))

for(t in 1:T) {
  E[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.E)
}
tau.E ~ dgamma(2,0.2)
sigma.E <- sqrt(1/tau.E)

for(t in 1:T) {
  for(i in 1:N.occ[t]) {
    occ.obs[ind.occ[i,t],t] ~ dbin(occ[ind.occ[i,t],t], 30)
  }
}

for (i in 1:N) {
  logit(occ[i,1]) <- x.occ[i,1]
  x.occ[i,1] <- mu.occ[1] + c.occ*U[i ]
}
for(t in 2:T) {
  for (i in 1:N) {
    logit(occ[i,t]) <- x.occ[i,t]
    x.occ[i,t] <- mu.occ[t] + gamma.occ*(x.occ[i,t-1] -mu.occ[t-1]) + U[i + N*(t-1)]
  }
}

for(t in 1:T) {
  mu.occ[t] ~ dnorm(0,tau.mu.occ)
}
tau.mu.occ ~ dgamma(0.1,0.02)

gamma.occ ~ dunif(-1,1)
c.occ <- 1/sqrt(1 - pow(gamma.occ , 2))

for(t in 1:T) {
  U[(1 + (t-1)*N) : (N + (t-1)*N)] ~car.normal(adj[], weights[], num[], tau.U)
}
tau.U ~ dgamma(2,0.2)
sigma.U <- sqrt(1/tau.U)
}

```

## B Trace plots

To show that I have a somewhat good sample of the target distribution, I show trace plots for model 4. Results from the other models were similar.

### B.1 Convergence

Figure 15 shows trace plot for the first iterations for  $\gamma^p$ ,  $\gamma^\phi$  and  $\sigma_U$  which shows convergence. Similar plots were made for other parameters, but they seemed to have good starting values and convergence was not possible to spot. The marginal distributions seem to reach values much earlier than 1 000 iterations, which is my burn-in.

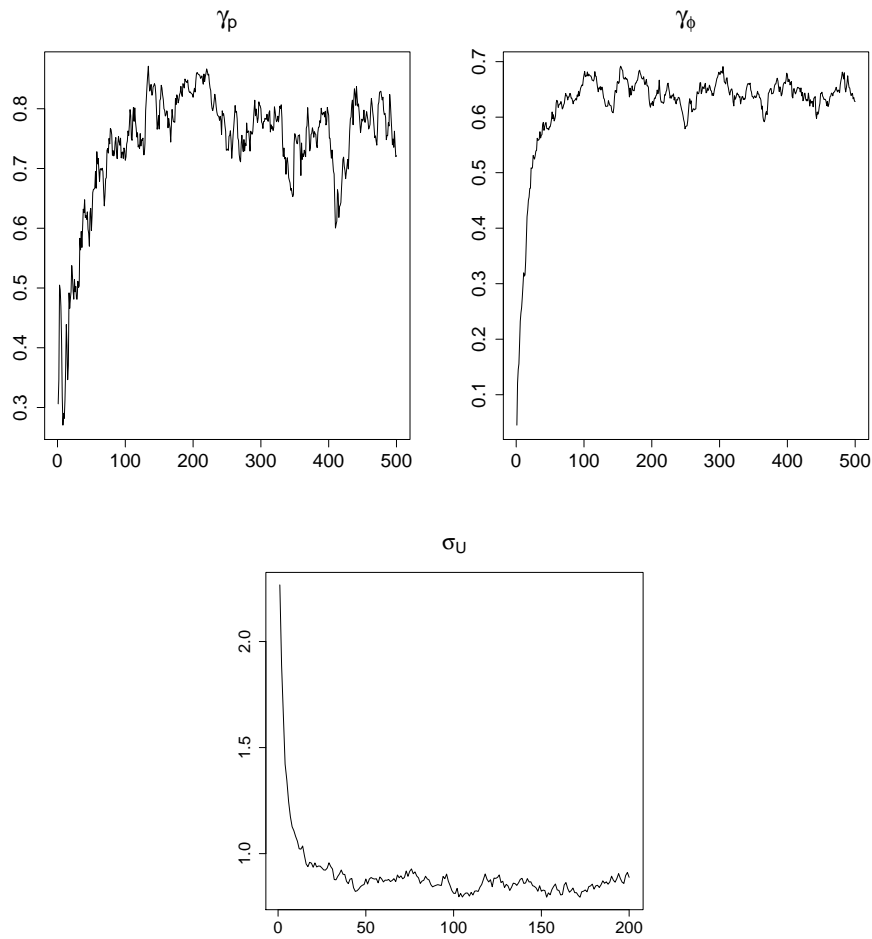


Figure 15: This figure shows trace plots of the first iterations for  $\gamma^p$ ,  $\gamma^\phi$  and  $\sigma_U$ .

## B.2 Stability

The trace plots shows that bigger samples for  $\beta_1$  and  $\sigma_E$  would be preferable, but for parameters of which I focus my analysis  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$  and  $\gamma_p$  the simulations seem stable.

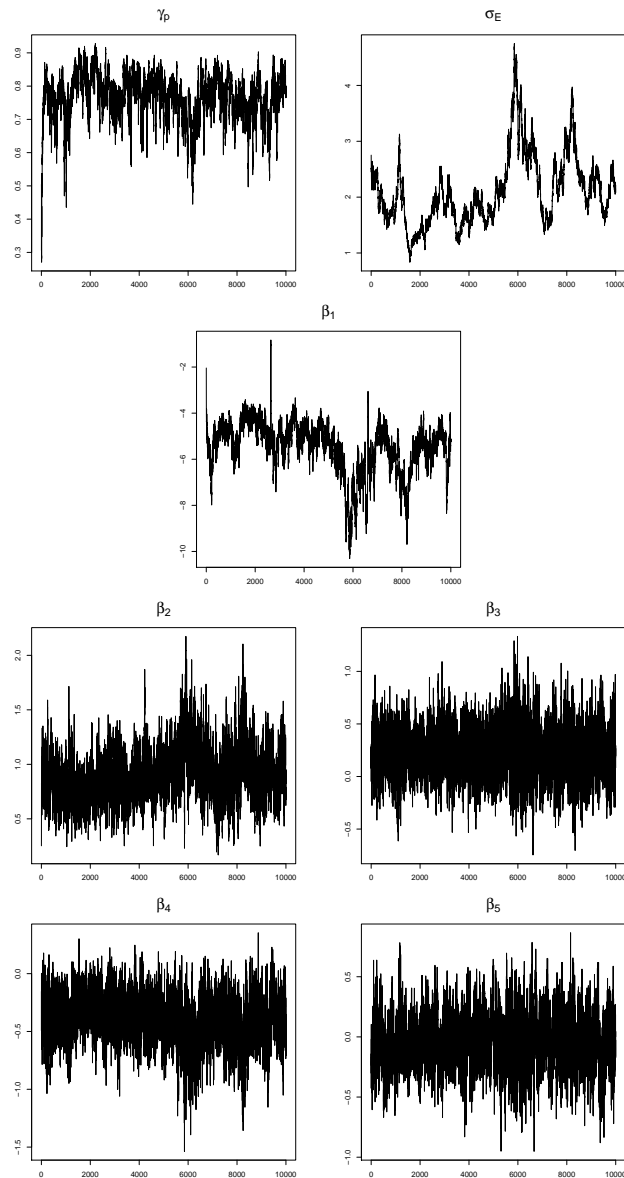


Figure 16: Trace plots of all parameters of interest in model 4.

## C Model 0a

For model 0a, I tested several priors. For these priors, the DIC results were very different. For this model the expectation for  $p_{i,t}^*$  is defined as

$$\mu_{i,t}^p = \mu_t^p \sim N(0, \tau_{\mu^p})$$

The differences on the tested model 0a's are the priors used for  $\tau_{\mu^p}$ :

Model	Definition of $\tau_{\mu^p}$
i	$\tau_{\mu^p} \sim \text{dgamma}(0.1, 0.02)$
ii	$\tau_{\mu^p} = 0.1$
iii	$\tau_{\mu^p} = 0.05$
iv	$\tau_{\mu^p} = 0.01$

Model	$\bar{D}$	$\hat{D}$	pD	DIC
i	373.307	230.560	142.747	516.054
ii	586.587	435.585	151.002	737.590
iii	491.386	356.430	134.956	626.342
iv	432.985	287.010	145.974	578.959

As shown in figure 17, the differences on simulated  $\mu_t^p$  are visible.

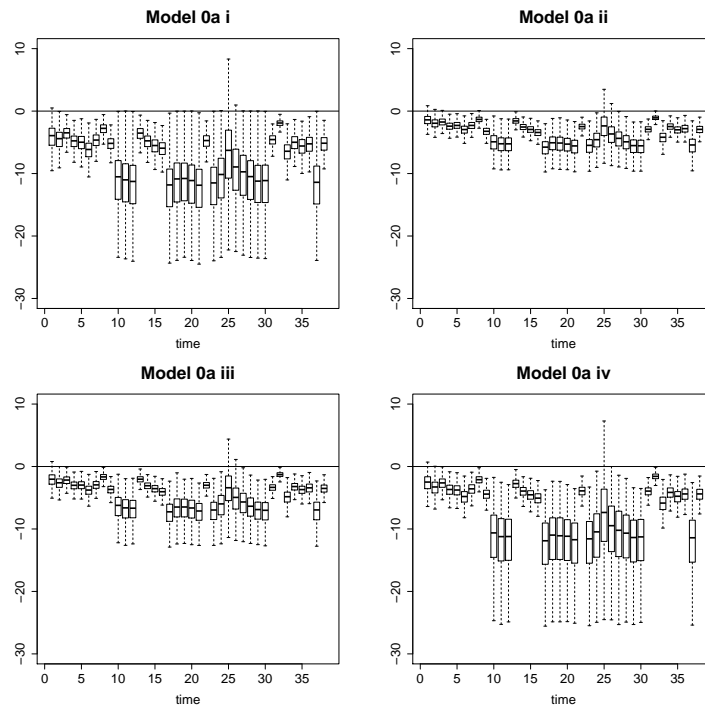


Figure 17: This figure shows box plots for  $\mu_t^p$  for the models 0a i, ii, iii and iv.

## D Adjustment for t=1 in AR(1) processes

To make the process stationary, I made the following adjustment for the first time point,  $t = 1$ .

$$p_{i,1}^* = \mu_{i,1}^p + c_p \mathbf{E}_{i,1}, \quad c_p = \frac{1}{\sqrt{1-\gamma_p^2}}$$

For  $t \geq 2$ :

$$p_{i,t}^* = \mu_{i,t}^p + \gamma_p * (p_{i,t-1}^* - \mu_{i,t-1}^p) + \mathbf{E}_{i,t-1}$$

These calculations show equal variance for all  $t$ .

$$\begin{aligned} \text{Var}[\mathbf{E}_t] &= \mathbf{\Sigma} \\ \text{Var}[\mathbf{p}_1^*] &= c_p^2 \text{Var}[\mathbf{E}_1] \\ &= \frac{1}{1-\gamma_p^2} \mathbf{\Sigma} \\ \text{Var}[\mathbf{p}_2^*] &= \gamma_p^2 \text{Var}[\mathbf{p}_1^*] + \text{Var}[\mathbf{E}_2] \\ &= \gamma_p^2 \frac{1}{1-\gamma_p^2} \mathbf{\Sigma} + \frac{1-\gamma_p^2}{1-\gamma_p^2} \mathbf{\Sigma} \\ &= \frac{\gamma_p^2 \mathbf{\Sigma} + (1-\gamma_p^2) \mathbf{\Sigma}}{1-\gamma_p^2} \\ &= \frac{1}{1-\gamma_p^2} \mathbf{\Sigma} \end{aligned}$$

For  $t \geq 2$  in general:

$$\begin{aligned} \text{Var}[\mathbf{p}_t^*] &= \gamma_p^2 \text{Var}[\mathbf{p}_{t-1}^*] + \text{Var}[\mathbf{E}_t] \\ &= \gamma_p^2 \frac{1}{1-\gamma_p^2} \mathbf{\Sigma} + \frac{1-\gamma_p^2}{1-\gamma_p^2} \mathbf{\Sigma} \\ &= \frac{\gamma_p^2 \mathbf{\Sigma} + (1-\gamma_p^2) \mathbf{\Sigma}}{1-\gamma_p^2} \\ &= \frac{1}{1-\gamma_p^2} \mathbf{\Sigma} \end{aligned}$$

The restriction of  $c_p = \frac{1}{\sqrt{1-\gamma_p^2}}$  was not possible to use for model 0a and 0b, However, simulated  $c_p$  was not very different from this.

## References

- L. Heier, G.O. Storvik, S.A. Davis, H. Viljugrein, V.S. Ageyev, E. Klassovskaya, and N.C. Stenseth. Emergence, spread, persistence and fade-out of sylvatic plague in kazakhstan. *Proceedings of the Royal Society B: Biological Sciences*, 2011.
- J.K. Kruschke and K. John. *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Academic Press, 2010.
- K. Linné Kausrud, H. Viljugrein, A. Frigessi, M. Begon, S. Davis, H. Leirs, V. Dubyanskiy, and N.C. Stenseth. Climatically driven synchrony of gerbil populations allows large-scale plague outbreaks. *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):1963, 2007.
- D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. *OpenBUGS*, 2011. URL <http://www.openbugs.info/w/>.
- D.J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. *WinBUGS 1.4.3*, 2007. URL <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, 2004.
- N.I. Samia, K.S. Chan, and N.C. Stenseth. A generalized threshold mixed model for analyzing nonnormal nonlinear time series, with application to plague in kazakhstan. *Biometrika*, 94(1):101, 2007.
- N.C. Stenseth, N.I. Samia, H. Viljugrein, K.L. Kausrud, M. Begon, S. Davis, H. Leirs, VM Dubyanskiy, J. Esper, V.S. Ageyev, et al. Plague dynamics are driven by climate variation. *Proceedings of the National Academy of Sciences*, 103(35):13110, 2006.