Modeling Regional Disease Spread Over Time Using a Dynamic Spatio-temporal Model – With an Application to Porcine Epidemic Diarrhea Virus data in Iowa, U.S.

J. Ji
Iowa State University

Chong Wang
Iowa State University, chwang@iastate.edu

M. Rotolo
Iowa State University

Jeffrey J. Zimmerman
Iowa State University, jjzimm@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/vdpam_pubs

Part of the Large or Food Animal and Equine Medicine Commons, Veterinary Infectious Diseases Commons, and the Veterinary Preventive Medicine, Epidemiology, and Public Health Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/vdpam_pubs/193. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Veterinary Diagnostic and Production Animal Medicine at Iowa State University Digital Repository. It has been accepted for inclusion in Veterinary Diagnostic and Production Animal Medicine Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Modeling Regional Disease Spread Over Time Using a Dynamic Spatio-temporal Model – With an Application to Porcine Epidemic Diarrhea Virus data in Iowa, U.S.

Abstract
Regional surveillance is important for detecting the incursion of new pathogens and informing disease monitoring and control programs. Modeling disease distribution over time can provide insight into the development of more efficient regional surveillance approaches. Herein we propose a Bayesian spatio-temporal model to describe the distribution of porcine epidemic diarrhea virus (PEDV) in Iowa USA. Model parameters are estimated through a Bayesian spatio-temporal model approach which can account for missing values. For illustration, we apply the proposed model to PEDV test results from the Iowa State University Veterinary Diagnostic Laboratory (ISU-VDL). A simulation study carried out to evaluate the model showed that the proposed model captured the pattern of PEDV distribution and its spatio-temporal dependence.

Keywords
PEDV, Spatio-temporal model, Bayesian analysis

Disciplines
Large or Food Animal and Equine Medicine | Veterinary Infectious Diseases | Veterinary Preventive Medicine, Epidemiology, and Public Health

Comments
This is a manuscript of an article published as Ji, J., C. Wang, M. Rotolo, and J. Zimmerman. "Modeling Regional Disease Spread Over Time Using a Dynamic Spatio-temporal Model–With an Application to Porcine Epidemic Diarrhea Virus data in Iowa, US." Preventive Veterinary Medicine (2020): 105053. DOI: 10.1016/j.prevetmed.2020.105053. Posted with permission.

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.
Modeling Regional Disease Spread Over Time Using a Dynamic Spatio-temporal Model – With an Application to Porcine Epidemic Diarrhea Virus data in Iowa, U.S.

J. Ji\textsuperscript{a}, C. Wang\textsuperscript{a,b}, M. Rotolo\textsuperscript{b}, J. Zimmerman\textsuperscript{b}

\textsuperscript{a}Department of Statistics, College of Liberal Arts and Sciences, Iowa State University, Ames, Iowa, 50011, Corresponding author: Ju Ji, jji@iastate.edu
\textsuperscript{b}Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, 50011

Abstract

Regional surveillance is important for detecting the incursion of new pathogens and informing disease monitoring and control programs. Modeling disease distribution over time can provide insight into the development of more efficient regional surveillance approaches. Herein we propose a Bayesian spatio-temporal model to describe the distribution of porcine epidemic diarrhea virus (PEDV) in Iowa USA. Model parameters are estimated through a Bayesian spatio-temporal model approach which can account for missing values. For illustration, we apply the proposed model to PEDV test results from the Iowa State University Veterinary Diagnostic Laboratory (ISU-VDL). A simulation study carried out to evaluate the model showed that the proposed model captured the pattern of PEDV distribution and its spatio-temporal dependence.

Keywords: PEDV, spatio-temporal model, Bayesian analysis

1. Introduction

The aim of this project was to explore area surveillance methods for livestock farms, with the long-term objective of developing timely and practical surveillance procedures for detecting the introduction of emerging and/or foreign animal diseases (FAD) into a region. Time-to-detection is a critical feature in disease control because delay allows the pathogen to spread and, thereby, increases the difficulty and cost of achieving control and/or elimination. For example, in a study of livestock farms in California, Carpenter et al. (2011) estimated that a 7 day delay in the detection of foot-and-mouth disease virus (FMDV) would result in a median of 13 infected premises holding...
677 animals, whereas a 22 day delay would result in a median of 745 infected premises with 6,211 animals. However, the imperative of rapid detection must be balanced by the cost of surveillance. An extreme example is the case of bovine spongiform encephalopathy for which the cost of finding one positive case in slaughter cattle in the European Union was reported as €2.3 million for the year 2003 (Heim and Mumford 2005). Essentially, the challenge is efficient disease detection at an affordable cost.

As reviewed by Saif et al. (2019), porcine epidemic diarrhea virus (PEDV) is an RNA virus in family Coronaviridae causing enteric infection in pigs and capable of producing extensive mortality in PEDV-susceptible neonatal piglets. Clinical outbreaks of PED were first reported in 1971 in England, but PEDV roused little attention until catastrophic outbreaks were reported in China, Thailand, and Korea beginning in 2007. Previously free of the pathogen, PEDV was detected in the U.S. in 2013 (Stevenson et al. 2013) and is estimated to have caused the deaths of 8 million piglets and economic losses of $481 to $929 million (USD) in 2014 (Paarlberg 2014). Immediately after its detection, efforts to diagnose PEDV infections in commercial swine herds resulted in the submission of massive numbers of samples for testing to the Iowa State University Veterinary Diagnostic Laboratory (ISU-VDL). Testing results, when combined with precise location data, presented a unique opportunity to evaluate spatial surveillance approaches for the detection of a pathogen introduced into a totally naive population using routine diagnostic testing data. Pragmatically, the use of routine diagnostic submissions in on-going surveillance offers both extensive cost savings (sample collection cost is avoided) and the opportunity for real-time detection.

Typically, contagious infectious diseases spread widely following their initial introduction, reach a peak prevalence, and then establish a cycle in the population that mirrors changes in population immunity over time. To study the spread of infectious diseases, several spatio-temporal Bayesian models have been proposed and evaluated in the literature (Denis et al. 2018; Knorr-Held 2000; Richardson et al. 2006; Lawson 2013; Watson et al. 2017). In most of these studies, the goal is to identify areas with high or low disease prevalence and to forecast disease prevalence. For example, a Bayesian spatio-temporal conditional autoregressive model was proposed to analyze and forecast the prevalence of antibodies to Borrelia burgdorferi in domestic dogs in the U.S. (Watson et al. 2017). Bayesian methods provide great advantage in the analysis of complex models with complicated data structure. Given the flexibility and generality of the Bayesian framework, we are allowed to cope with complex problems, especially when we have latent variables, missing data, or multilayered
probability specifications ([Gelman et al. 2013](#)). Notably, unlike maximum likelihood estimation in the frequentist inference, Bayesian methods have the advantage of avoiding integration over the latent variables or missing values, especially given the complexity of the model there is usually no closed form of the integration. The computation is easier if using Markov chain Monte Carlo (MCMC) methods to draw samples from the posterior distribution ([Besag and Mondal 2013](#) [West and Harrison 2006](#) Chap.15). This is relevant to the present study because we propose to use a dynamic model with latent variables and about 27.5% of PEDV test result data from the ISU-VDL are missing.

Iowa is a highly agricultural state located in the north-central U.S. The state is 150,930 square kilometers in size and divided into 99 counties of relatively uniform size ([Figure 1](#)). Federal Orders (June 5, 2014, January 4, 2016) entitled “Reporting, Herd Monitoring and Management of Novel Swine Enteric Coronavirus Diseases” required that swine producers report cases of swine coronavirus infections and provide premises identification number, date of sample collection, type of unit sampled, the diagnostic test used, and the results of testing. This provided an extensive dataset on PEDV in Iowa swine farms. The initial dataset included a total of 30,843 PEDV PCR test results from samples collected from premises located in the state of Iowa and tested at the ISU-VDL between May 2014 and March 2017. A data query was performed to retrieve all PEDV test results. In order to maintain data integrity, the results were processed to remove any result that did not have an associated valid address representing a swine premises. Premises identification numbers (PIN) and premises submission level identifiers were entered into Google Earth and if a swine premises was located at the entered address the result was maintained. If the address or PIN provided did not pertain to a swine premises, the result was removed from the data set. Binary (positive/non-positive) results were used for summary and analyses. In our study, we selected the 6-month period, August 2016 to January 2017, with the most complete data to address the spread of PEDV.

In this paper, we propose a Bayesian spatio-temporal model for the distribution of PEDV at the county level in Iowa that accounts for the effect of spatial distance among counties. As illustrated in Figure [1](#) the observed proportion of PEDV in each county from August 2016 to January 2017 shows a clear spatio-temporal dependence. There are a few counties with no tested samples at certain time points. The top two reasons that can address the majority of missing are farmers may not submit samples for testing if the disease status is not a concern and the farm lost all of the
animals due to acute outbreak. There might be a small portion of missing due to reasons such as producers may have elected to submit samples for testing to a laboratory other than the ISU-VDL. Our goal is to build a model reflects the major trend of missing. Thus to account for the missing values, we added constraints based on the two major reasons described earlier to the Bayesian spatio-temporal model.

The paper is organized as follows: Section 2 presents the proposed spatio-temporal model and the corresponding Bayesian inference including MCMC algorithm and choice of prior. Section 3 shows the simulation results. Section 4 reports the results of PEDV data analysis. Section 5 provides the discussion of the results and evaluation of the model performance. Section 6 concludes the paper with a summary and some future works.

2. Modeling procedure

2.1. Bayesian spatio-temporal model for PEDV distribution

In this section, we propose a Bayesian spatio-temporal model to describe changes in the distribution of PEDV over time. Suppose there are $n$ counties in our study, and the time $t$ ranges from 0 to $T$ with each time point representing one month. Let $N_{s,t}$ be the total number of samples submitted for testing from county $s$ at time $t$, and $y_{s,t}$ is the number of samples tested positive, $s \in \{1, 2, ..., n\}$ and $t \in \{0, 1, ..., T\}$. $y_{s,t}$ follows a binomial distribution:

$$y_{s,t} \sim \text{Binomial}(N_{s,t}, p_{s,t}),$$

where $p_{s,t}$ is the probability of samples in county $s$ at time $t$ being tested PEDV positive, and $\xi_{s,t}$ is its logit transformation. Here $\{\xi_{s,t}\}$ are latent variables which have spatio-temporal dependence. That is, $\xi_{s,t}$ is influenced by its previous time value $\xi_{s,t-1}$ and the status of neighbor counties of $s$ at time $t-1$. To account for this spatio-temporal dependence among counties, we model the increment of $\xi_{s,t}$, $\Delta \xi_{s,t} = \xi_{s,t} - \xi_{s,t-1}$, to be a linear function of the probabilities of being tested as PEDV positive from other counties. In addition, the magnitude of the impact of neighbor counties on county $s$ is also related to their distance. Therefore, we propose the model in the following way:
\[ \Delta \xi_{s,t} = \beta_0 + \alpha \sum_{s \neq s'} \exp(-d_{s,s'})p_{s',t-1} + \phi_{s,t}. \]

Here \( \beta_0 \) is the intercept which accounts for the average increment of \( \xi \) across time without considering any spatial interactions. \( d_{s,s'} \) is the distance between county \( s \) and county \( s' \), and \( p_{s',t-1} \) is the probability of samples from county \( s' \) at time \( t-1 \) being tested PEDV positive. In this model, the spatial impact of county \( s' \) on county \( s \) is quantified as \( \alpha \sum_{s \neq s'} \exp(-d_{s,s'})p_{s',t-1} \), which decreases exponentially as the distance \( d_{s,s'} \) increases. Thus for any county, its neighbor counties have larger influence than other counties that are far away. The parameter \( \alpha \) controls the overall magnitude of the spatial impacts on \( \xi_{s,t} \) from all other counties \( s' \neq s \) at the previous time point.

The structure of the model is autoregressive and thus \( \xi \) has the Markov property. There are also some other covariates that potentially affect \( \Delta \xi_{s,t} \), such as the number of farms in the county and the transportation infrastructure (interstate, US and state highways) in the county. The latter was considered because pigs are routinely moved between farms and counties in the production cycle. The presence of more farms and more developed transportation infrastructure may have affected the distribution of the virus. Thus, we include covariates in the model:

\[ \xi_{s,t} = \xi_{s,t-1} + X_{s,t}' \beta + \alpha \sum_{s \neq s'} \exp(-d_{s,s'})p_{s',t-1} + \phi_{s,t}, \]  

for \( t = 1, 2, ..., T \) and \( s = 1, 2, ..., n \).

Here \( X_{s,t} = (1, X_{1,s,t}, X_{2,s,t}, ..., X_{p,s,t})' \) is the \( p+1 \) dimensional vector of constant one and the covariates that affect the distribution of PEDV. \( \beta = (\beta_0, \beta_1, ..., \beta_p)' \) is the corresponding vector of regression parameters. \( \phi_{s,t} \)'s are the random errors accounting for other unmeasured factors, which are independently identically distributed as \( N(0, \tau^2) \). The initial states, \( \xi_{s,0} \), for \( s = 1, 2, ..., n \), are modeled as \( \xi_{s,0} \overset{i.i.d.}{\sim} N(\mu, \sigma^2) \).

2.2. Considering missing values

Within the considered time period, there are some counties that have no samples submitted from farms for testing in at least one month, i.e., missing values. Presumably, farms do not submit samples for testing if there are no clinical signs of disease among piglets. That is, we assume that farms submit samples for testing only if the proportion of clinically-affected piglets is relatively
high. Then one assumption we can make in the analysis is that the decision not to submit samples
means that the disease status at that time point is not a concern, thus a low proportion (less than
0.1) of PEDV can be assumed. Therefore, if a county $s$ at time $t$ has $y_{s,t}$ missing, we put constraints
on $\xi_{s,t}$ as $\xi_{s,t} \in [\text{logit}(0.001), \text{logit}(0.1)]$ to make the proportion of PEDV to be less than 0.1.

In addition, if a farm has submitted samples at time $t - 1$ but none at time $t$, i.e. $y_{s,t-1}$ is observed but $y_{s,t}$ is missing, it is very likely that all the piglets in that farm died. PEDV infection remains endemic, and it is impossible for the infection to disappear naturally within a short period of time. In many cases during the acute outbreak, farms were unable to produce healthy piglets for weeks or even a few months. These cases do not fit in our model, since the model is only intended to describe the spatio-temporal trends in PEDV. Therefore, the spatio-temporal dependency will be cut for such cases, which results the corresponding $\xi_{s,t}$ to be no longer influenced by others and we propose a noninformative distribution on $[\text{logit}(0.001), \text{logit}(0.1)]$ for $\xi_{s,t}$, i.e. $\xi_{s,t} \sim \text{Unif}(\text{logit}(0.001), \text{logit}(0.1))$.

To incorporate these constraints into the model, we first define the following five sets:

$$O = \{(s, t) : \text{for all } y_{s,t} \text{ that are observed}\}$$
$$M = \{(s, t) : \text{for all } y_{s,t} \text{ that are missing}\}$$
$$\Omega = \{(s, t) : t \geq 1\}$$
$$\Omega_1 = \{(s, t) : t \geq 1, y_{s,t} \text{ is missing at } (s, t) \text{ but observed at } (s, t-1)\}$$
$$\Omega_2 = \Omega \setminus \Omega_1.$$ 

For $t = 0$ and all county $s = 1, 2, ..., n$, we have $\xi_{s,0} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, which is the same as the model described in section 2.1. For $t = 1, 2, ..., T$, and county $s = 1, 2, ..., n$, we have the following model:

$$(s, t) \in \Omega_1 \quad \text{Unif}(\text{logit}(0.001), \text{logit}(0.1)), \quad (2)$$
$$(s, t) \in \Omega_2 \quad \xi_{s,t} = \xi_{s,t-1} + \bm{X}_{s,t}'\bm{\beta} + \alpha \sum_{s' \neq s'} \exp(-d_{s,s'})p_{s',t-1} + \phi_{s,t}.$$ 

For $(s, t) \in \Omega_1$, this model sets density of $\xi_{s,t}$ as a constant, which means its distribution is
noninformative. For \((s, t) \in \Omega_2\), the model is the same as the Bayesian spatio-temporal model shown in equation [1].

2.3. Joint probability density function (pdf) of the PEDV outcome

Here we develop the joint pdf of the logit transformed probability of tested positive \(\xi = (\xi_0, \xi_1, ..., \xi_T)\), where \(\xi_t = (\xi_{1,t}, ... \xi_{n,t})\) represents a vector of logit transformed probability of tested positive at time \(t\) for all counties. The joint pdf of \(\xi\) can be derived based on the spatio-temporal model described above. Given that there are two different model settings, one is without missing value described in section 2.1 and the other is under missing values with constraints described in section 2.2 we have two different joint pdf forms.

2.3.1. Joint pdf under complete data setting

For \(t = 0\), given \(\xi_{0,t}\)'s are independently and identically distributed, \(\xi_0 = (\xi_{1,0}, ..., \xi_{n,0})'\) has the density as:

\[
f(\xi_0|\mu, \sigma^2) \propto (\sigma^2)^{-n/2} \prod_{s=1}^{n} \exp\left( -\frac{(\xi_{s,0} - \mu)^2}{2\sigma^2} \right).
\]

For \(t = 1, ..., T\), based on the Markov property, the joint density of \(\xi_{1:T} = (\xi_1, ..., \xi_T)\) has the following form

\[
f(\xi_{1:T}|\xi_0, \beta, \alpha, \tau^2) = f(\xi_1|\xi_0, \beta, \alpha, \tau^2) ... f(\xi_T|\xi_{T-1}, \beta, \alpha, \tau^2)
\]

\[
\propto (\tau^2)^{-nT} \prod_{s=1}^{n} \prod_{t=1}^{T} \exp\left( -\frac{1}{2\tau^2} (\xi_{s,t} - \xi_{s,t-1} - X_{s,t}^\prime \beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'}p_{s',t-1}))^2 \right).
\]

Thus the joint pdf for \(\xi\) is derived as:
\[ f(\xi | \beta, \alpha, \tau^2, \sigma^2) = f(\xi_0 | \mu, \sigma^2) f(\xi_{1:T} | \xi_0, \beta, \alpha, \tau^2) \]  

(3)

\[ \propto (\sigma^2)^{-n/2} \prod_{s=1}^{n} \exp \left( -\frac{\left( \xi_{s,0} - \mu \right)^2}{2\sigma^2} \right) \times \]

\[ (\tau^2)^{-\frac{nT}{2}} \prod_{s=1}^{n} \prod_{t=1}^{T} \exp \left( -\frac{1}{2\tau^2} (\xi_{s,t} - \xi_{s,t-1} - X'_{s,t} \beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',t-1})^2 \right). \]

Let \( y = (y_1, y_2, ..., y_T) \) be the observed outcomes for all counties at all time, where each \( y_t = (y_{1,t}, y_{2,t}, ..., y_{n,t})' \) is a vector of observed outcome for all counties at time point \( t \).

Finally, the joint density of \( y \) and \( \xi \) has the following form:

\[ f(y, \xi | \beta, \alpha, \mu, \tau^2, \sigma^2) = f(y | \xi) f(\xi | \beta, \alpha, \mu, \tau^2, \sigma^2), \]  

(4)

with \( f(y | \xi) = \prod_{s=1}^{n} \prod_{t=1}^{T} \exp(\xi_{s,t})^{y_{s,t}} [1 + \exp(\xi_{s,t})]^{-N_{s,t}} \) and \( f(\xi | \beta, \alpha, \tau^2, \sigma^2) \) as shown in equation (3).

2.3.2 Joint pdf under missing value with constraints setting

The joint pdf of \( \xi \) is then derived as:

\[ f(\xi | \beta, \alpha, \tau^2, \sigma^2) = f(\xi_0 | \mu, \sigma^2) f(\xi_{1:T} | \xi_0, \beta, \alpha, \tau^2) \]  

(5)

\[ \propto (\sigma^2)^{-n/2} \prod_{s=1}^{n} \exp \left( -\frac{\left( \xi_{s,0} - \mu \right)^2}{2\sigma^2} \right) \prod_{(s,t) \in M} I(logit(0.001) \leq \xi_{s,t} \leq logit(0.1)) \times \]

\[ (\tau^2)^{-\frac{|\Omega_2|}{2}} \prod_{(s,t) \in \Omega_2} \exp \left( -\frac{1}{2\tau^2} (\xi_{s,t} - \xi_{s,t-1} - X'_{s,t} \beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',t-1})^2 \right), \]

where \( |\Omega_2| \) is the number of elements in set \( \Omega_2 \) and \( I() \) is an indicator function.

Finally, the joint density of \( y \) and \( \xi \) has the following form:
\[ f(y, \xi | \beta, \alpha, \mu, \tau^2, \sigma^2) = f(y|\xi) f(\xi | \beta, \alpha, \mu, \tau^2, \sigma^2), \]  
\[ \text{with } f(y|\xi) = \prod_{(s,t) \in O} \left[ \exp(\xi_{s,t}) \right]^{y_{s,t}} \left[ 1 + \exp(\xi_{s,t}) \right]^{-N_{s,t}}. \]  

2.4. Bayesian inference

Bayesian posterior inference will be used to estimate the parameters of interest include \( \beta = (\beta_0, \beta_1, ..., \beta_p)' \), \( \alpha \), \( \mu \), \( \sigma^2 \) and \( \tau^2 \). The prior distribution for each of the parameters are as follows:

- \( \beta_k \sim N(0, s_{\beta}^2) \) independently for each \( k = 0, 1, ..., p \) \hspace{1cm} (7)
- \( \alpha \sim N(0, s_\alpha^2) \) \hspace{1cm} (8)
- \( \mu \sim N(0, s_\mu^2) \) \hspace{1cm} (9)
- \( \tau^{-2} \sim \text{Gamma}(a, b) \) \hspace{1cm} (10)
- \( \sigma^{-2} \sim \text{Gamma}(c, d) \). \hspace{1cm} (11)

The parameterization of the Gamma distribution prior for \( \tau^{-2} \) has the mean of \( \frac{a}{b} \) and variance \( \frac{a}{b^2} \), and \( \sigma^{-2} \) has the mean of \( \frac{c}{d} \) and variance \( \frac{c}{d^2} \). The conjugate non-informative priors are used for all model parameters, where \( s_\beta, s_\alpha \) and \( s_\mu \) take the value 100, and \( a = b = c = d = 0.01 \). Sensitivity analysis has been performed for different choices of prior parameters \( (a, b, c, d = 0.1, 0.01, 0.001, \text{or} 0.0001) \). The results showed that different values of prior distribution parameters only have little impact on the estimates. Given the joint pdf shown in equations (4) and (6), the joint posterior distributions can be derived as:

\[ p(\xi, \beta, \alpha, \mu, \tau^2, \sigma^2 | y) \propto f(y, \xi | \beta, \alpha, \mu, \tau^2, \sigma^2) f(\beta) f(\alpha) f(\mu) f(\tau^2) f(\sigma^2). \]  

Markov chain Monte Carlo (MCMC) with a combination of Gibbs and Metropolis-Hastings (MH) steps is used to draw samples from the posterior distributions of the parameters sequentially. Based on the draws from the posterior distributions we can make inferences for the model.
parameters. The detailed posterior sampling algorithm is described in section Appendix: MCMC Algorithm and the computing program is written in R.

The parameters $\beta$, $\alpha$, $\mu$, $\tau^2$ and $\sigma^2$ have closed form of full conditional posterior distributions from which we can directly sample. However, the conditional posterior distribution of $\boldsymbol{\xi}$ given other parameters does not have a closed form. We use random walk MH to update $\xi_{s,t}$ one by one, which means the proposal distribution is $N(\xi_{s,t}^{(m-1)}, \delta^2_{s,t})$. Here, how to scale $\delta_{s,t}$ is of importance: too small values will make the chain move very slowly; too large values will result in very low acceptance rate of the proposals. To choose $\delta_{s,t}$ in an appropriate way, we use the idea of adaptive MCMC in [Haario et al. (2001)]. At the beginning, we use pre-specified $\delta_{s,t}$ such as $\delta_{s,t} = 0.2$, to carry out a MCMC simulation. After an initial period, for each $(s,t)$, a good choice of $\delta_{s,t}$ would be the standard derivation of the posterior samples of $\xi_{s,t}$ in this MCMC. After that, a new MCMC is implemented with these new $\delta_{s,t}$’s.

3. Simulation Study

A simulation study was performed to evaluate the performance of both models (complete and missing values). Under each model, there are two different settings. In both settings we set the number of counties and the total period of time to be the same as the real data. In the first setting, we have the number of samples tested from each county at each time, $N_{s,t}$, fixed as 25 for all counties at all times, since the averaged number of samples tested in the PEDV test data over all counties is about 25. In the second setting, we fix $N_{s,t}$ as 100 for all counties at all times. Posterior means of parameters are used as the true parameters in the simulation. Each setting has 100 simulations. For each simulation, the data is generated from the true parameters, and then we applied the proposed models to estimate the parameters. The parameter estimations from the 100 simulations are compared with the true parameters to evaluate the models performance.

In the simulation under complete data model, we first generate $\xi_{s,0}$, for $s = 1, \ldots, 31$, from $N(\mu, \sigma^2)$, where $\mu = -3.9164$ and $\sigma^2 = 2.0935$. Then for $t = 1, \ldots, 5$, the $\xi_t$ is generated as equation (1), where model parameters are set to be the posterior means from real data analysis, i.e. $\beta_0 = -0.0942$, $\alpha = 0.9705$, and $\tau^2 = 2.1679$. Given the generated $\xi_{s,t}$, the number of samples tested positive, $y_{s,t}$, can then be generated from a binomial distribution given size of $N_{s,t} = 25$ or 100 and probability of $p_{s,t} = \frac{\exp(\xi_{s,t})}{1 + \exp(\xi_{s,t})}$. The result based on 100 simulation for model with complete values is shown in Table 1.
In the simulation under missing data model, we first generate $\xi_{s,0}$, for $s = 1, ..., 85$, from $N(\mu, \sigma^2)$, where $\mu = -3.7945$ and $\sigma^2 = 0.9721$. Then for $t = 1, ..., 5$, $\xi_t$ is generated as equation (1), where model parameters are set to be the posterior means from real data analysis, i.e. $\beta_0 = -0.7324$, $\alpha = 0.9188$, and $\tau^2 = 3.0639$. Given that we assume the proportion of PEDV for samples that are missing in the PEDV data are less than 0.1 for all $s = 1, ..., 85$ and $t = 0, 1, ..., 5$. We calculate the proportion of missing among samples with proportion of PEDV less than 0.1 in the PEDV data as $P\% = \frac{\text{number of missing samples}}{\text{number of missing samples} + \text{number of observed samples with observed proportion of PEDV} \leq 0.1}$. In the simulated data for $(s,t)$'s with values of $\xi_{s,t} \leq \text{logit}(0.1)$, we randomly set $P\%$ of them as missing. The simulation result for model with missing values constraints is shown in Table 2.

4. Results of the PEDV Data

The proposed spatio-temporal model is applied to the PEDV test results from the ISU-VDL with testing dates from August 2016 to January 2017. The testing results are at the county level to protect client confidentiality. We are interested in modeling the monthly spatio-temporal pattern of PEDV. In Figure 1, we can see that the averaged observed proportion of PEDV over all counties is very low, with an average of 0.023. Over time, the proportion of PEDV continues to increase until the end of the sixth month. Samples for testing were received monthly from 85 out of 99 counties during this 6-month period. We proposed two models, one is for the complete data setting which will be applied in section 4.1, and the other is for the incomplete data setting and will be applied in section 4.2.

4.1. Results under complete data setting

After deleting the counties with missing values to obtain a complete dataset, only 31 counties are left. The number of samples submitted for testing from each county at each month ranges from 1 to 331. We have considered six covariates in the model in order to achieve a more accurate estimation of the spatio-temporal PEDV trend. These six covariates are season, number of interstate highways, US highways, and state highways in each county, area of each county in square miles and the number of farms in each county divided by one thousand. However, none of these six covariates has significant contribution to the model (Table 3). Thus, our final model only includes the intercept, $\beta_0$. 

11
To start the MCMC iteration, a common choice of the initial value of $\xi_{s,t}$ is the logit of the observed proportion of PEDV $\logit(\hat{p}_{s,t})$ with $\hat{p}_{s,t} = y_{s,t}/N_{s,t}$. However, given the number of samples tested in some counties are small (less than 3), it is easy to have a $y_{s,t}$ taking value of 0 or $N_{s,t}$, which makes the corresponding $\hat{p}_{s,t}$ to be either 0 or 1. In such case, $\logit(\hat{p}_{s,t})$ is no longer valid. To avoid setting $p_{s,t}$ as an arbitrary number such as 0.001 or 0.999, we borrowed the idea of Agresti and Coull (1998) and used the following weighted average to approximate $p_{s,t}$:

$$\tilde{p}_{s,t} = \hat{p}_{s,t} \frac{N_{s,t}}{N_{s,t} + z^2_\alpha} + 0.5 \frac{z^2_\alpha}{N_{s,t} + z^2_\alpha},$$

where $z_\alpha$ is the $\alpha$ quantile of the standard normal. Here we choose $z_{0.975} \approx 1.96$. This makes $\tilde{p}_{s,t}$ fall into $(0,1)$ and shrinks the observed proportion of PEDV to 0.5. When $N_{s,t}$ goes up, the weight assigned to $\hat{p}_{s,t}$ increases and the shrinkage becomes less. We implement this correction for all $p_{s,t}$’s, and then use $\logit(\tilde{p}_{s,t})$ as the initial value of $\xi_{s,t}$.

The estimation of parameters were based on 15,000 MCMC iterations after disregarding the first 10,000 burn-in period. The Gelman-Rubin diagnostic test was performed to check the convergence, and we ran three different chains with starting values as -50, 0, and 50. The scale reduction factor from the Gelman-Rubin diagnostic test is 1 means that between variance and within chain variances are equal, which is an indication of convergence. The multivariate effective sample sizes is 3581 for all parameters of interest, based on 15,000 MCMC with 10,000 burn-in. Table 4 shows the posterior means, posterior standard deviations and 95% Bayesian credible intervals (CI) from the model without missing values.

4.2. Model under missing value with constraints setting

The spatio-temporal model under missing value with constraints setting is applied to the same PEDV dataset. We are keeping all testing results from those 85 counties regardless if they have any values of missing or not during the 6 months period. Notice that there are 99 counties in total in Iowa, but farms tend to be concentrated in certain counties (farms are not evenly distributed across counties) and not all farms will have had clinical PEDV during this six-month period. Among the 85 counties over the six-month period, there are about 27.5% values missing. This missing rate is calculated as dividing the total number of samples that are observed by $85 \times 6 = 510$. Figure 2 shows the observed proportion of PEDV for the 85 counties that with at least one observation over 6 months, where each line represents a time series for one county. Again, similar as under the
complete data setting, we first considered six covariates in the model but none of these covariates have significant contribution to the model (Table 5). The final model only includes the intercept \( \beta_0 \) with result shown in Table 6.

5. Discussion

The result of the simulation study with complete values is shown in Table 1, and the simulation result for model with missing values constraints is shown in Table 2. We can see for each parameter the average of the 100 posterior means from the 100 simulations is close to the true value, compared with the standard error of these 100 posterior means, for both models. Both simulation results support the validity of the proposed models under complete data setting and missing data settings, respectively.

For the PEDV data analysis result for complete data model with only the intercept, Table 4 shows the posterior means, posterior standard deviations and 95% Bayesian credible intervals (CI) from the model without missing values. Except the intercept all of the parameters are significant, given their 95% credible intervals do not cover 0. \( \alpha > 0 \) indicates that the proportion of PEDV of a county is influenced by other county \( s' \), where \( s' \neq s \). If the distance \( d_{s,s'} \) between two counties is larger, the influence on \( \xi_{s,t} \) will be smaller. Also, if the probability of tested positive, \( p_{s',t-1} \), of county \( s' \) is smaller, its influence on \( \xi_{s,t} \) will be smaller. The result for missing data model with only the intercept is shown in Table 6. The parameter estimates are very similar to the model with complete data except that the estimates for \( \beta_0 \) is smaller and has significant contribution to the model (with 95% Bayesian credible interval doesn’t cover 0). This means that without considering any spatial interactions, the average increment of the proportion of PEDV across time under the missing value model is larger than under the complete data model. Also notice that the posterior standard deviation for the parameters for model with missing values (Table 6) are all smaller than those without missing values (Table 4). Together with the consistent posterior mean estimates between the two models, it is shown that the model with missing values provide more precise estimates by incorporating more data. Thus when there are missing values, the analysis with missing value data model is preferred to the analyses with complete data only.
6. Conclusions and Future Work

In this project, we used a spatio-temporal model to account for the changes in the spatio-temporal pattern of PEDV. The PEDV test result data we used were derived from routine diagnostic testing, which presents a huge advantage in terms of lower cost and timeliness (real-time surveillance). Applying the proposed model to PEDV test results from the ISU-VDL, the parameters estimation results showed strong evidence of distance-dependency at a regional (county) level. Given the importance of efficient regional surveillance, the proposed model can be used to predict the changes in the distribution of PEDV over time and thus lead to better guidelines for detecting the incursion of new pathogens and informing disease monitoring and control programs. For future work, the proposed model could inform the development of sampling guidelines that could be more timely and sensitive in detecting emerging diseases than surveillance based on simple random sampling or passive methods.

Acknowledgments

This project was funded from this grant: Zimmerman J, Wang C, Main R. An exploration of surveillance based on spatial sampling using ISU VDL submissions. Swine Health Information Center (shic@swinehealth.org).
7. References

References

Agresti, A., Coull, B.A., 1998. Approximate is better than “exact” for interval estimation of binomial proportions. The American Statistician 52, 119–126.

Besag, J., Mondal, D., 2013. Exact goodness-of-fit tests for markov chains. Biometrics 69, 488–496.

Carpenter, T.E., O’Brien, J.M., Hagerman, A.D., McCarl, B.A., 2011. Epidemic and economic impacts of delayed detection of foot-and-mouth disease: a case study of a simulated outbreak in california. Journal of Veterinary Diagnostic Investigation 23, 26–33.

Denis, M., Cochard, B., Syahputra, I., De Franqueville, H., Tisné, S., 2018. Evaluation of spatio-temporal bayesian models for the spread of infectious diseases in oil palm. Spatial and Spatio-temporal Epidemiology 24, 63–74.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. Chapman and Hall/CRC.

Haario, H., Saksman, E., Tamminen, J., et al., 2001. An adaptive metropolis algorithm. Bernoulli 7, 223–242.

Heim, D., Mumford, E., 2005. The future of bse from the global perspective. Meat Science 70, 555–562.

Knorr-Held, L., 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine 19, 2555–2567.

Lawson, A.B., 2013. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. Chapman and Hall/CRC.

Paarlberg, P., 2014. Updated estimated economic welfare impacts of porcine epidemic diarrhea virus (PEDV). Technical Report.

Richardson, S., Abellan, J.J., Best, N., 2006. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in yorkshire (uk). Statistical Methods in Medical Research 15, 385–407.
Saif, L.J., Wang, Q., Vlasova, A.N., Jung, K., Xiao, S., 2019. Coronaviruses. Diseases of Swine, 488–523.

Stevenson, G.W., Hoang, H., Schwartz, K.J., Burrough, E.R., Sun, D., Madson, D., Cooper, V.L., Pillatzki, A., Gauger, P., Schmitt, B.J., et al., 2013. Emergence of porcine epidemic diarrhea virus in the united states: clinical signs, lesions, and viral genomic sequences. Journal of Veterinary Diagnostic Investigation 25, 649–654.

Watson, S.C., Liu, Y., Lund, R.B., Gettings, J.R., Nordone, S.K., McMahan, C.S., Yabsley, M.J., 2017. A bayesian spatio-temporal model for forecasting the prevalence of antibodies to borrelia burgdorferi, causative agent of lyme disease, in domestic dogs within the contiguous united states. PloS One 12, e0174428.

West, M., Harrison, J., 2006. Bayesian forecasting and dynamic models. Springer Science & Business Media.
Appendix: Figures and Tables

Figure 1: Monthly observed proportion of PEDV among counties in Iowa from Aug. 2016 to Jan. 2017 in the Veterinary Diagnostic Laboratory at Iowa State University (ISU-VDL) testing result.
Figure 2: Plot of observed proportion of PEDV over 6 months with missing values

Table 1: Average and standard error of posterior means for simulation studies of the model with complete data (based on 100 simulations)

| $N_{s,t}$ | parameter | true value | averaged posterior mean | SE of posterior mean |
|-----------|-----------|------------|-------------------------|----------------------|
| 25        | $\beta_0$ | -0.0942    | -0.0933                 | 0.1490               |
|           | $\alpha$  | 0.9705     | 0.9589                  | 0.0907               |
|           | $\mu$     | -3.9164    | -3.9675                 | 0.3851               |
|           | $\sigma^2$| 2.0935     | 2.3269                  | 1.0320               |
|           | $\tau^2$  | 2.1679     | 2.3425                  | 0.4663               |
| 100       | $\beta_0$ | -0.0942    | -0.0907                 | 0.1299               |
|           | $\alpha$  | 0.9705     | 0.9690                  | 0.0710               |
|           | $\mu$     | -3.9164    | -3.9858                 | 0.2980               |
|           | $\sigma^2$| 2.0935     | 2.2916                  | 0.8586               |
|           | $\tau^2$  | 2.1679     | 2.2044                  | 0.3194               |
Table 2: Average and standard error of posterior means for simulation studies of the model with missing values.
(based on 100 simulations)

| $N_{s,t}$ | parameter | true value | averaged posterior mean | SE of posterior mean |
|-----------|-----------|------------|-------------------------|----------------------|
| 25        | $\beta_0$ | -0.7324    | -0.6220                 | 0.1433               |
|           | $\alpha$ | 0.9188     | 0.9065                  | 0.0176               |
|           | $\mu$    | -3.7945    | -3.7387                 | 0.1665               |
|           | $\sigma^2$ | 0.9721    | 0.7542                  | 0.3117               |
|           | $\tau^2$ | 3.0639     | 3.1134                  | 0.4806               |
| 100       | $\beta_0$ | -0.7324    | -0.6425                 | 0.1248               |
|           | $\alpha$ | 0.9188     | 0.9073                  | 0.0161               |
|           | $\mu$    | -3.7945    | -3.8084                 | 0.1345               |
|           | $\sigma^2$ | 0.9721    | 0.8769                  | 0.2261               |
|           | $\tau^2$ | 3.0639     | 3.1577                  | 0.3665               |

Table 3: Posterior results for complete data model with all six covariates analyzed with PEDV data

| parameter | posterior mean | posterior SD | LB 95% CI | UB 95% CI |
|-----------|----------------|--------------|-----------|-----------|
| $\beta_0$ | -0.7775        | 0.9739       | -2.6839   | 1.1574    |
| $\beta_{I-HW}$ | -0.0036   | 0.2970       | -0.5853   | 0.5800    |
| $\beta_{US-HW}$ | -0.0643   | 0.2191       | -0.4938   | 0.3640    |
| $\beta_{ST-HW}$ | -0.0390   | 0.0994       | -0.2348   | 0.1534    |
| $\beta_{area}$ | 0.0013     | 0.0020       | -0.0027   | 0.0053    |
| $\beta_{farmnum}$ | 1.1584    | 1.1540       | -1.0956   | 3.4993    |
| $\beta_{season}$ | -0.0491   | 0.4094       | -0.8427   | 0.7807    |
| $\alpha$ | 1.0039        | 0.2859       | 0.4458    | 1.5909    |
| $\mu$    | -4.1299       | 0.5359       | -5.2853   | -3.1673   |
| $\sigma^2$ | 2.4397      | 0.6305       | 1.4174    | 3.8664    |
| $\tau^2$ | 2.4416        | 1.3383       | 0.7108    | 5.8476    |

Table 4: Posterior results for complete data model without covariates analyzed with PEDV data

| parameter | posterior mean | posterior SD | LB 95% CI | UB 95% CI |
|-----------|----------------|--------------|-----------|-----------|
| $\beta_0$ | -0.0942        | 0.1542       | -0.3959   | 0.2071    |
| $\alpha$ | 0.9705         | 0.2556       | 0.4748    | 1.4862    |
| $\mu$    | -3.9164        | 0.4285       | -4.7596   | -3.1277   |
| $\sigma^2$ | 2.0935      | 1.0108       | 0.7055    | 4.5333    |
| $\tau^2$ | 2.1679         | 0.5347       | 1.2651    | 3.3686    |
Table 5: Posterior results for missing values model with all six covariates analyzed with PEDV data

| Parameter | posterior mean | posterior SD | LB 95% CI | UB 95% CI |
|-----------|----------------|--------------|-----------|-----------|
| $\beta_0$ | -1.4744        | 0.5956       | -2.6594   | -0.3215   |
| $\beta_{I-HW}$ | 0.0172        | 0.1804       | -0.3342   | 0.3790    |
| $\beta_{US-HW}$ | -0.1283       | 0.1505       | -0.4328   | 0.1610    |
| $\beta_{ST-HW}$ | -0.0286       | 0.0685       | -0.1626   | 0.1052    |
| $\beta_{area}$ | 0.0019        | 0.0011       | -0.0004   | 0.0042    |
| $\beta_{farmnum}$ | 0.8997        | 0.8757       | -0.8077   | 2.6383    |
| $\beta_{season}$ | -0.4391       | 0.3368       | -1.1014   | 0.2153    |
| $\alpha$   | 0.9506         | 0.0856       | 0.7788    | 1.1141    |
| $\mu$      | -3.9780        | 0.2638       | -4.5710   | -3.5207   |
| $\sigma^2$ | 1.1178         | 0.3958       | 0.5244    | 2.0535    |
| $\tau^2$   | 2.9549         | 0.5472       | 2.0550    | 4.2167    |

Table 6: Posterior results for missing value model without covariates analyzed with PEDV data

| Parameter | posterior mean | posterior SD | LB 95% CI | UB 95% CI |
|-----------|----------------|--------------|-----------|-----------|
| $\beta_0$ | -0.7324        | 0.1338       | -1.0068   | -0.4806   |
| $\alpha$  | 0.9188         | 0.0886       | 0.7493    | 1.051     |
| $\mu$     | -3.7945        | 0.2272       | -4.5710   | -3.5207   |
| $\sigma^2$ | 0.9721         | 0.3640       | 0.4433    | 1.8664    |
| $\tau^2$  | 3.0639         | 0.5012       | 2.2247    | 4.1546    |
Appendix: MCMC Algorithm

For model under complete data setting

Given initial value $\alpha^{(0)}$, $\sigma^2(0)$, $\tau^2(0)$ and $\xi^{(0)}$, for iteration $m = 1,...M$:

1. Sample $\beta^{(m)}$ from a Multivariate Normal distribution with mean
   \[
   \left(\tau^{-2} \sum_{t=1}^{T} X_t^T X_t + s^{-2}_\beta I\right)^{-1} \left(\tau^{-2} \sum_{t=1}^{T} X_t^T (\xi_t - \xi_{t-1} - \alpha A p_{t-1})\right),
   \]
   and covariance matrix $\left(\tau^{-2} \sum_{t=1}^{T} X_t^T X_t + s^{-2}_\beta I\right)^{-1}$. Here $X_t = (X_{1,t},...,X_{n,t})'$ is a $n \times (p+1)$ matrix, and $A = (exp(d_{s,s'}))_{n \times n}$ is a square matrix with $(s,s')$ entry as $exp(-d_{s,s'})$ for all $s \neq s'$. $p_{t-1} = (p_{1,t-1},p_{2,t-1},...,p_{n,t-1})'$ is the vector of probabilities of tested positive for all counties at time $t - 1$.

2. Sample $\alpha^{(m)}$ from a normal distribution with mean $\frac{\sum_{s,t=1}^{n} \xi_{s,t} p_{s,t}}{\sum_{s,t=1}^{n} p_{s,t}}$, and variances $\frac{s^2 \sum_{s,t=1}^{n} (\xi_{s,t} - \xi_{s,0})^2}{\sum_{s,t=1}^{n} p_{s,t}}$.

3. Sample $\mu^{(m)} \sim N\left(\frac{\sum_{s,t=1}^{n} \xi_{s,t}}{\sum_{s,t=1}^{n} p_{s,t}}, \frac{s^2 \sum_{s,t=1}^{n} (\xi_{s,t} - \xi_{s,0}^*)}{\sum_{s,t=1}^{n} p_{s,t}}\right)$.

4. Use random walk MH to generate $\xi^{(m)}_{s,t}$: firstly generate a candidate $\xi_{s,t}^* \sim N(\xi_{s,t}^{(m-1)}, \delta_{s,t})$ and $U \sim Unif(0,1)$, and calculate the acceptance rate (AR). If $U$ is less than AR we accept the candidate $\xi_{s,t}^{(m)} = \xi_{s,t}^*$, otherwise we reject it and keep $\xi_{s,t}^{(m)} = \xi_{s,t}^{(m-1)}$. The acceptance rate for each $\xi_{s,t}$ is as follows.

   - For $t = 0$
     \[
     AR = \frac{(1 + \exp(\xi_{s,0}^*))^{-N_{s,0}} \exp(y_{s,0} \xi_{s,0}^*)}{(1 + \exp(\xi_{s,0}^{(m-1)}))^{-N_{s,0}} \exp(y_{s,0} \xi_{s,0}^{(m-1)})} \times \frac{\exp\left(-\frac{1}{2\tau^2} (\xi_{s,0}^* - \mu)^2\right)}{\exp\left(-\frac{1}{2\tau^2} (\xi_{s,0}^{(m-1)} - \mu)^2\right)} \times \frac{\exp\left(-\frac{1}{2\tau} \|\xi_{1} - \xi_{0}^* - X_1 \beta - \alpha A p_{0}\|^2\right)}{\exp\left(-\frac{1}{2\tau} \|\xi_{1} - \xi_{0}^{(m-1)} - X_1 \beta - \alpha A p_{0}^{(m-1)}\|^2\right)},
     \]

   - For $t = 1,\ldots,T - 1$
\[ AR = \frac{(1 + \exp(\xi^*_{s,t}))^{-N_{s,t}} \exp (y_{s,t} \xi^*_{s,t})}{(1 + \exp(\xi^{(m-1)}_{s,t}))^{-N_{s,t}} \exp (y_{s,t} \xi^{(m-1)}_{s,t})} \times \]
\[ \frac{\exp \left( -\frac{1}{2\tau^2} (\xi^*_{s,t} - \xi_{s,t-1} - \mathbf{X}'_{s,t}\beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',t-1})^2 \right)}{\exp \left( -\frac{1}{2\tau^2} (\xi^{(m-1)}_{s,t} - \xi_{s,t-1} - \mathbf{X}'_{s,t}\beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',t-1})^2 \right)} \times \]
\[ \frac{\exp \left( -\frac{1}{2\tau^2} \| \xi_{t+1} - \xi^*_{t} - \mathbf{X}_{t+1}\beta - \alpha \mathbf{A}_{t}(\mathbf{p}_{t+1}) \|^2 \right)}{\exp \left( -\frac{1}{2\tau^2} \| \xi_{t+1} - \xi^{(m-1)}_{t} - \mathbf{X}_{t+1}\beta - \alpha \mathbf{A}_{t}(\mathbf{p}_{t+1}) \|^2 \right)}, \]

- For \( t = T \)

\[ AR = \frac{(1 + \exp(\xi^*_{s,T}))^{-N_{s,T}} \exp (y_{s,T} \xi^*_{s,T})}{(1 + \exp(\xi^{(m-1)}_{s,T}))^{-N_{s,T}} \exp (y_{s,T} \xi^{(m-1)}_{s,T})} \times \]
\[ \frac{\exp \left( -\frac{1}{2\tau^2} (\xi^*_{s,T} - \xi_{s,T-1} - \mathbf{X}'_{s,T}\beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',T-1})^2 \right)}{\exp \left( -\frac{1}{2\tau^2} (\xi^{(m-1)}_{s,T} - \xi_{s,T-1} - \mathbf{X}'_{s,T}\beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',T-1})^2 \right)} \times \]
\[ \frac{\exp \left( -\frac{1}{2\tau^2} \| \xi_{t+1} - \xi^*_{t} - \mathbf{X}_{t+1}\beta - \alpha \mathbf{A}_{t}(\mathbf{p}_{t+1}) \|^2 \right)}{\exp \left( -\frac{1}{2\tau^2} \| \xi_{t+1} - \xi^{(m-1)}_{t} - \mathbf{X}_{t+1}\beta - \alpha \mathbf{A}_{t}(\mathbf{p}_{t+1}) \|^2 \right)}, \]

5. Sample \( \sigma^{2(m)} \) from the following inverse gamma distribution

\[ IG(c + \frac{n}{2}, d + \frac{1}{2} \sum_{s=1}^{n} (\xi_{s,0} - \mu)^2). \]

6. Sample \( \tau^{2(m)} \) from the following inverse gamma distribution

\[ IG \left( \frac{nT}{2} + a, b + \frac{1}{2} \sum_{s=1}^{n} \sum_{t=1}^{T} (\xi_{s,t} - \xi_{s,t-1} - \mathbf{X}'_{s,t}\beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'})p_{s',t-1})^2 \right). \]

For model under missing values with constraints setting

For iteration \( m = 1, \ldots, M \):

1. Sample \( \mathbf{B}^{(m)} \) from a Multivariate Normal distribution with mean
\[
\left( \tau^{-2} \sum_{(s,t) \in \Omega_2} X_{s,t}X_{s,t}^T + s_\beta^{-2} I \right)^{-1} \left[ \frac{1}{\tau^2} \sum_{(s,t) \in \Omega_2} \left( \xi_{s,t} - \xi_{s,t-1} - \alpha \sum_{s' \neq s} \exp(-d_{s,s'}) p_{s',t-1} \right) X_{s,t} \right]
\]

and variance \[\left[ \tau^{-2} \sum_{(s,t) \in \Omega_2} X_{s,t}X_{s,t}^T + s_\beta^{-2} I \right]^{-1}.\]

2. Sample \(\alpha^{(m)}\) from a normal distribution with mean

\[
s_\alpha^2 \sum_{(s,t) \in \Omega_2} \left( \sum_{s' \neq s} \exp(-d_{s,s'}) (\xi_{s,t} - \xi_{s,t-1} - X'_{s,t} \beta) \right)
\]

and variance \[s_\alpha^2 \sum_{(s,t) \in \Omega_2} \left[ \sum_{s' \neq s} \exp(-d_{s,s'}) p_{s',t-1} \right]^2 + \tau^2\]

3. Sample \(\mu^{(m)}\) similarly as in complete case, only change the acceptance rates as follows. To simplify the notation, for any vector \(\nu\) we set \(\|\nu\|^2_{t \in J} = \sum_{i \in J} v_i^2\). Further we set \(I_{s,t}\) to be the indicator of missingness of \(y_{s,t}\):

\[
I_{s,t} = \begin{cases} 
1, & \text{if } y_{s,t} \text{ is observed,} \\
0, & \text{if } y_{s,t} \text{ is missing.} 
\end{cases}
\]

- For \(t = 0\)

\[
AR = \left[ \frac{(1 + \exp(\xi^*_{s,0}))^{-N_{s,0}} \exp(y_{s,0} \xi^*_{s,0})}{(1 + \exp(\xi^{(m-1)}_{s,0}))^{-N_{s,0}} \exp(y_{s,0} (\xi^{(m-1)}_{s,0}))} \right]^{I_{s,0}} \times \frac{\exp(-\frac{1}{2\tau^2} (\xi^*_{s,0} - \mu)^2)}{\exp(-\frac{1}{2\tau^2} (\xi^{(m-1)}_{s,0} - \mu)^2)} \times
\]

\[
\exp \left( -\frac{1}{2\tau^2} \|\xi_1 - \xi^*_{0} - X_1 \beta - \alpha A p^*_{0}(l_{(i,1)} \in \Omega_2) \|^2 \right) \quad \exp \left( -\frac{1}{2\tau^2} \|\xi^{(m-1)}_1 - X_1 \beta - \alpha A p^{(m-1)}_{0}(l_{(i,1)} \in \Omega_2) \|^2 \right)
\]

- For \(t = 1, ..., T - 1\)

23
\[
AR = \left[ \frac{(1 + \exp(\xi_{s,t}^*))^{-N_{s,t}} \exp(y_{s,t} \xi_{s,t}^*)}{(1 + \exp(\xi_{s,t}^{(m-1)})^{-N_{s,t}} \exp(y_{s,t} \xi_{s,t}^{(m-1)})} \right]^{I_{s,t}} \times \\
\exp \left( -I((s,t) \in \Omega_2) \times \frac{1}{2\pi^2}(\xi_{s,t}^* - \xi_{s,t-1} - X_{s,t}' \beta - \alpha \sum_{s \neq s'} \exp(-d_{s,s'}p_{s',t-1})^2) \right) \\
\exp \left( -I((s,t) \in \Omega_2) \times \frac{1}{2\pi^2}(\xi_{s,t}^{(m-1)} - \xi_{s,t-1} - X_{s,t}' \beta - \alpha \sum_{s \neq s'} \exp(-d_{s,s'}p_{s',t-1})^2) \right) \\
\exp \left( -\frac{1}{2\pi^2} \| \xi_{t+1} - \xi_{t}^* - X_{t+1} \beta - \alpha Ap_{i}^{(m-1)} \|_{(i,t+1) \in \Omega_2}^2 \right) \\
\exp \left( -\frac{1}{2\pi^2} \| \xi_{t+1} - \xi_{t}^{(m-1)} - X_{t+1} \beta - \alpha Ap_{i} \|_{(i,t+1) \in \Omega_2}^2 \right) \\
\]

- For \( t = T \)

\[
AR = \left[ \frac{(1 + \exp(\xi_{s,T}^*))^{-N_{s,T}} \exp(y_{s,T} \xi_{s,T}^*)}{(1 + \exp(\xi_{s,T}^{(m-1)})^{-N_{s,T}} \exp(y_{s,T} \xi_{s,T}^{(m-1)})} \right]^{I_{s,T}} \times \\
\exp \left( -I((s,T) \in \Omega_2) \times \frac{1}{2\pi^2}(\xi_{s,T}^* - \xi_{s,T-1} - X_{s,T}' \beta - \alpha \sum_{s \neq s'} \exp(-d_{s,s'}p_{s',T-1})^2) \right) \\
\exp \left( -I((s,T) \in \Omega_2) \times \frac{1}{2\pi^2}(\xi_{s,T}^{(m-1)} - \xi_{s,T-1} - X_{s,T}' \beta - \alpha \sum_{s \neq s'} \exp(-d_{s,s'}p_{s',T-1})^2) \right) \\
\]

5. Sample \( \sigma^2 \sim IG \left( c + \frac{1}{2}, d + \frac{1}{2} \sum_{s=1}^{m} (\xi_{s,0} - \mu)^2 \right) \).

6. Sample \( \tau^2 \sim IG \left( \frac{\Omega_1}{2} + a, b + \frac{1}{2} \sum_{(s,t) \in \Omega_1} (\xi_{s,t} - \xi_{s,t-1} - X_{s,t}' \beta - \alpha \sum_{s' \neq s} \exp(-d_{s,s'}p_{s',t-1})^2 \right) \).