Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Bayesian Penalized Spline Models for the Analysis of Spatio-Temporal Count Data

Cici Chen[a]*, Jon Wakefield[b], Håvard Rue[c], Steve Self[d], Zijian Feng[e] and Yu Wang[e]

Contents

[a] *Department of Biostatistics, Brown University*
[b] *Department of Statistics, University of Washington*
[c] *Norwegian University of Science and Technology*
[d] *Fred Hutchinson Cancer Research Center*
[e] *Chinese Center for Disease Control and Prevention*
* *Correspondence to: Department of Biostatistics, Brown University. E-meail: Cici_Bauer@brown.edu*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

1. China Central North Analysis

In Figure 1 we show the central north region, in relation to the whole of China.

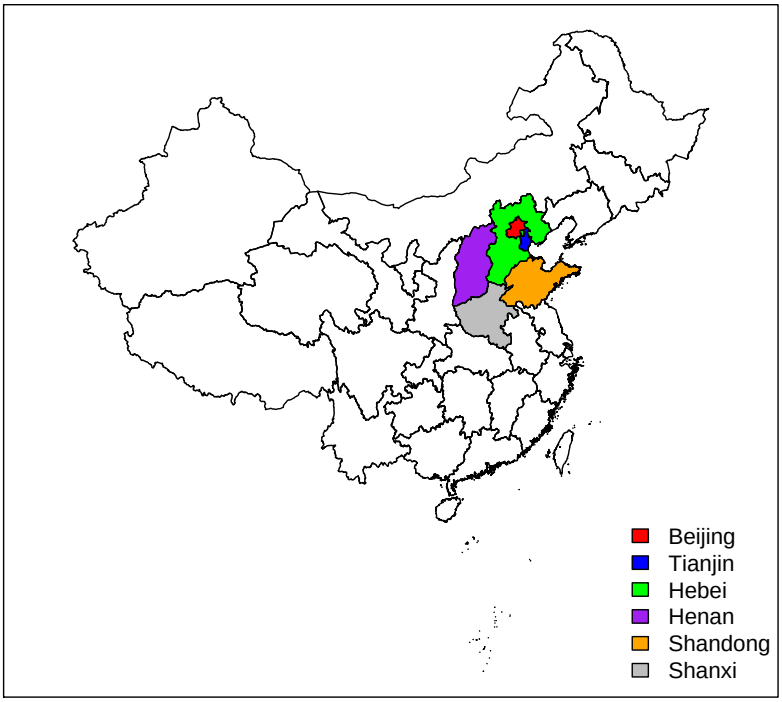


Figure 1. The central north region, in relation to China as a whole.

A movie of the log SMRs and the smoothed estimates

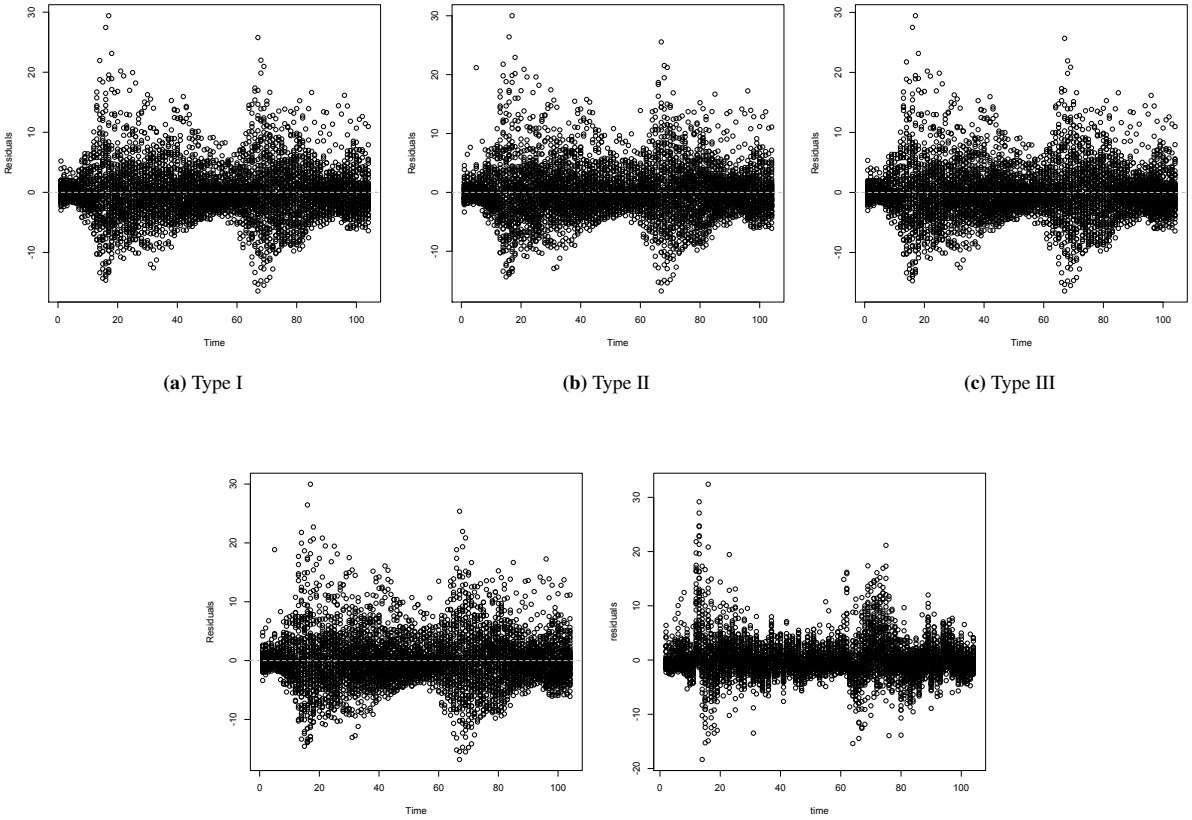
$$\hat{\alpha} + \widehat{\gamma}_t + \sum_{k=1}^{K} \widehat{b}_{kt} B_{ik}$$

can be found at http://www.stat.brown.edu/cbauer/ along with R code for fitting the models described in the paper.

We use standardized residuals which are defined as

$$r_{it} = \frac{y_{it} - \widehat{y}_{it}}{\sqrt{\widehat{y}_{it}}},$$

$i = 1, \ldots, I$, $t = 1, \ldots, T$. These are plotted against time t, for the Type I-IV models in Figure 2. We see that the fit is generally good, though we slightly underestimate the observed counts at the start of the epidemic.



(a) Type I **(b)** Type II **(c)** Type III

(d) Type IV **(e)** Epidemic/endemic Model

Figure 2. Residuals versus time, for the Type I - IV models, compared to the epidemic/endemic model in (1) .

2. Gaussian Markov Random Field (GMRF) Models

Second-Order Random Walk Model

In the paper we use GMRF models in both time and space, so here we give some technical details on these models, beginning with the second-order random walk (RW2) model. We write $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_T]^{\mathrm{T}} \sim \mathrm{RW2}(\tau_\gamma^{-1})$. This model corresponds to an improper Gaussian distribution of the form

$$\pi(\boldsymbol{\gamma}|\tau_\gamma) \propto {\tau_\gamma}^{(T-2)/2} \exp\Big(-\frac{\tau_\gamma}{2} \boldsymbol{\gamma}^T \boldsymbol{A}_{\mathrm{T}} \boldsymbol{\gamma} \Big),$$

with the structure matrix $\boldsymbol{A}_{\text{T}}$ given by

$$\boldsymbol{A}_{\text{T}} = \begin{bmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{bmatrix}, \tag{1}$$

with the blank space corresponding to zeros. Note that $\boldsymbol{A}_{\text{T}}$ has rank $T-2$. It is natural to model dependencies in a conditional fashion. Let σ_γ^2 denote the variance of γ_t, $\sigma_\gamma^2 = 1/\tau_\gamma$. The conditional mean and variance are:

$$\begin{aligned} \text{E}[\gamma_t|\boldsymbol{\gamma}_{-t}, \sigma_\gamma^2) &= \frac{4}{6}(\gamma_{t+1} + \gamma_{t-1}) - \frac{1}{6}(\gamma_{t+2} + \gamma_{t-2}) \\ \text{var}(\gamma_t|\boldsymbol{\gamma}_{-t}, \sigma_\gamma^2) &= \frac{\sigma_\gamma^2}{6} \end{aligned} \tag{2}$$

for $2 < t < n-2$. Some intuition into the conditional mean form may be gained by defining

$$p(t) = \beta_0 + \beta_1 t + \beta_2 t^2/2.$$

If we carry out a least squares fit through the points

$$(t-2, \gamma_{t-2}), \quad (t-1, \gamma_{t-1}), \quad (t+1, \gamma_{t+1}), \quad (t+2, \gamma_{t+2}).$$

Then $\text{E}[\gamma_t|\boldsymbol{\gamma}_{-t}, \sigma_\gamma^2]$ from (2) corresponds to $\widehat{p}(t)$. Hence, the RW2 model is often described as a *local quadratic smoother*. If we fix the second-order random walk at $1, \ldots, T$ then

$$\begin{aligned} \text{E}[\gamma_{T+k}|\gamma_1, \ldots, \gamma_T, \sigma_\gamma^2] &= (1+k)\gamma_T - k\gamma_{T-1} \\ \text{var}(\eta_{T+k}|\gamma_1, \ldots, \gamma_T, \sigma_\gamma^2) &= \sigma_\gamma^2(1 + 2^2 + \cdots + k^2). \end{aligned}$$

Therefore the conditional mean is a linear extrapolation based on the last two points and the variance is cubic, since

$$1 + 2^2 + \cdots + k^2 = k(k+1)(2k+1)/6.$$

This explains why long term predictive explode under this model, though the locally quadratic mean form results in less bias than the random walk first-order model.

The model has rank $T-2$ with the rank 2 deficiency indicating that the RW2 distribution is invariant to the addition of any line $\alpha + \boldsymbol{\gamma t}$ to $\boldsymbol{\gamma}$. We have an intercept in our model and to make the temporal random effects γ_t identifiable, we put both $\sum_t \gamma_t = 0$ and $\sum_t t\gamma_t = 0$ constraints on $\boldsymbol{\gamma}$.

Intrinsic CAR Model

We now describe the intrinsic conditional autoregressive (ICAR) model introduced by (2). This model has been widely adopted in disease mapping context. Letting $\boldsymbol{u} = [u_1 \ldots, u_N]$ represent the collection of spatial random effects, the

conditional distributions are

$$u_i|u_j, j \neq i \sim \mathrm{N}\left(\frac{1}{m_i}\sum_{j\sim i} u_j, \frac{1}{m_i\tau_u}\right), \tag{3}$$

where m_i is the number of neighbors of area i and τ_u is the conditional precision. The notation $j \sim i$ indicates that areas j and i are neighbors. A common approach to define neighbors is if two areas share boundaries. Other neighborhood schemes are possible however, for example, (3) defines the neighborhood structure as a function of the distance between centroids.

An alternative way to write the ICAR "density" of $\boldsymbol{u}$ is to use the joint distribution

$$\pi(\boldsymbol{u}|\tau_u) \propto {\tau_u}^{(I-1)/2} \exp\left[-\frac{\tau_u}{2}\sum_{i\sim j}(u_i - u_j)^2\right]. \tag{4}$$

Notice that this distribution is only defined by the pairwise difference of the elements of $\boldsymbol{u}$ and so the overall level is unspecified, and hence we do not have a proper density. It is easy to see that the ICAR model is an example of a Gaussian Markov Random Field (GMRF) by rewriting the joint distribution of $\boldsymbol{u}$ in terms of the precision matrix $\boldsymbol{A}_{\mathrm{S}}$, see (4)

$$\pi(\boldsymbol{u}|\tau_u) \propto {\tau_u}^{(I-1)/2} \exp\left[-\frac{\tau_u}{2}\boldsymbol{u}^T\boldsymbol{A}_{\mathrm{S}}\boldsymbol{u}\right], \tag{5}$$

where the structure matrix $\boldsymbol{A}_{\mathrm{S}}$ has entries

$$A_{\mathrm{S}_{ij}} = \begin{cases} m_i & i = j, \\ -1 & i \sim j, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Here we have assumed that each of the areas is connected to at least one other. Note that the structure matrix $\boldsymbol{A}_{\mathrm{S}}$ has rank $I-1$ because the sum of each row (or column) is zero, again confirming that the ICAR prior distribution used for the spatial random effect $\boldsymbol{u}$ is improper. To make the random effects u_i identifiable in the posterior distribution, we need to either exclude the intercept, or if an intercept is included in the model with a flat prior then we place a sum-to-zero constraint on the spatial random effects. Both approaches reduce the number of parameters by 1, which is required for identifiability. The second approach was recommended in (2) and we follow their recommendation when we use the ICAR prior in the paper.

Kroneker Forms for the Interaction Models

In this section we describe in more detail the form of the joint distribution of the interaction terms $\boldsymbol{b}$. All four types can be represented as GMRF models with the general form

$$\pi(\boldsymbol{b}|\tau_b) \quad \propto \quad \exp\left[-\frac{\tau_b}{2}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}_{\mathrm{ST}}\boldsymbol{b}\right].$$

Let $\boldsymbol{A}_{\mathrm{S}}$ and $\boldsymbol{A}_{\mathrm{T}}$ denote the structure matrices in space and time respectively. The structure matrix of the spline coefficients is the Kronecker product of the two matrices, i.e. $\boldsymbol{A}_{\mathrm{ST}} = \boldsymbol{A}_{\mathrm{S}} \otimes \boldsymbol{A}_{\mathrm{T}}$ (5; 6).

For the Type I prior, since there is no spatial or temporal structure, $\boldsymbol{A}_{\mathrm{S}}$ and $\boldsymbol{A}_{\mathrm{T}}$ are both identity matrices with corresponding dimension giving the resulting structure matrix $\boldsymbol{A}_{\mathrm{ST}} = \boldsymbol{I}_{\mathrm{K}} \otimes \boldsymbol{I}_{\mathrm{T}}$, assuming K is the number of spline bases and T is the number of time points. For the Type II prior, which has temproal structure but no spatial structure we take $\boldsymbol{A}_{\mathrm{T}}$ of the form in (1), to get the structure matrix $\boldsymbol{A}_{\mathrm{ST}} = \boldsymbol{A}_{\mathrm{K}} \otimes \boldsymbol{I}_{\mathrm{T}}$. Note that in this case $\boldsymbol{A}_{\mathrm{ST}}$ has rank $K(T-2)$ because $\boldsymbol{A}_{\mathrm{T}}$ has rank $T-2$. Under the Type III prior, which has spatial structure only, we consider a first-order neighborhood structure for the basis functions and use the form (6) to obtain $\boldsymbol{A}_{\mathrm{ST}} = \boldsymbol{I}_{\mathrm{K}} \otimes \boldsymbol{A}_{\mathrm{T}}$, with rank $T(K-1)$. Finally, for the Type IV prior, we take both the structured spatial and temporal matrices to get $\boldsymbol{A}_{\mathrm{ST}} = \boldsymbol{A}_{\mathrm{K}} \otimes \boldsymbol{A}_{\mathrm{T}}$, with rank $(T-2)(K-1)$.

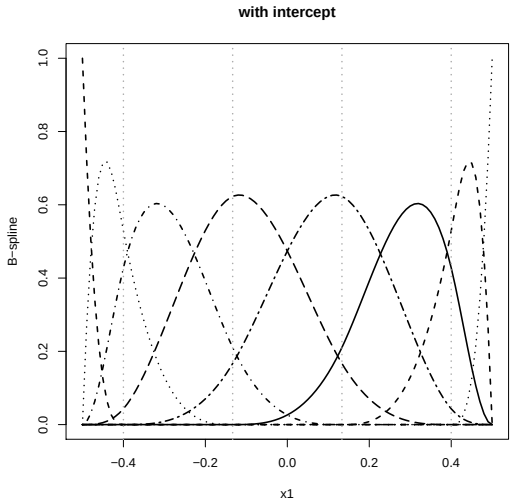


Figure 3. Cubic B spline basis functions in one dimension with $K = 4$ knots.

3. Tensor Product of B-Splines

There are many possibilities for spline modeling including natural thin plate splines (7), thin plate regression splines (which are the two-dimensional analog of natural cubic regression splines, see (8)) and P-splines (9) which use B-splines with a penalty on the differences. Radial bases have also been used in a spatial context (10). In our approach we use a tensor product of B-splines.

We first describe one-dimensional cubic B-splines in x

$$\sum_{k=1}^{K+4} \gamma_k B_k^3(x).$$

The total number of knots is K with knot locations κ_k, $k = 1, \ldots, K$. A B-spline with degree $d = 0$ is just a set of piecewise constant basis functions. Higher degree B-splines are constructed recursively from those of lower degree:

$$B_k^d(x) = \frac{x - \kappa_k}{\kappa_{k+d} - \kappa_k} B_k^{d-1}(x) + \frac{\kappa_{k+d+1} - x}{\kappa_{k+d+1} - \kappa_{k+1}} B_{k+1}^{d-1}(x).$$

This is equivalent to a truncated power series basis, but has more desirable numerical behavior. B-splines have many advantages (11). For example, with the are being bounded from above and for all x at most $d + 1$ bases are positive. Figure 3 shows cubic B spline basis functions in one dimension with $K = 4$ knots.

For two-dimensional surface fitting, the bivariate B-spline basis function is constructed as the tensor products of two univariate B-splines. Let $B_{k_1,1}(x_1)$, $k_1 = 1, ..., K_1$ and $B_{k_2,2}(x_2)$, $k_2 = 1, ..., K_2$ be cubic B-spline bases for x_1 and x_2 with $K_1 - 4$ and $K_2 - 4$ knots in each direction. Then the tensor product is

$$f(x_1, x_2) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} b_{k_1 k_2} B_{k_1 k_2}(x_1, x_2),$$

where

$$B_{k_1 k_2}(x_1, x_2) = B_{k_1,1}(x_1) B_{k_2,2}(x_2), \qquad k_1 = 1, ..., K_1; \;\; k_2 = 1, ..., K_2.$$

This is a low/intermediate rank representation. Figure 4 shows the bases functions in two dimensions with $K = 4$ knots in each direction. In the paper, for simplicity, we write the model as

$$\sum_{k=1}^{K} b_{kt} B_{ik}, \quad k = 1, \ldots, K,$$

where B_{ik} is a tensor product of cubic B-spline basis functions

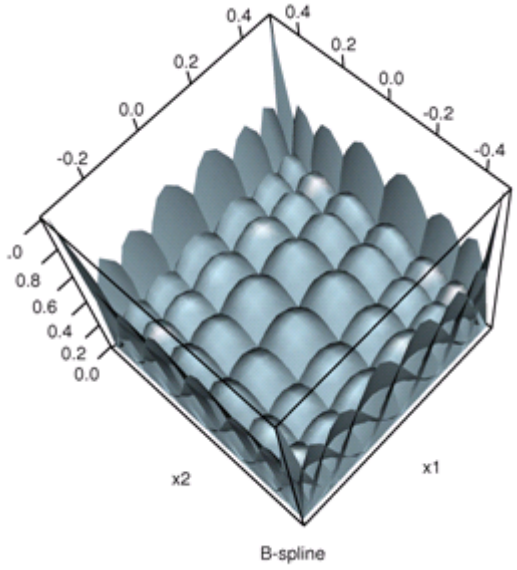


Figure 4. Tensor product of cubic B-splines with 4 knots in each direction.

4. Example code for WinBUGS and INLA

Example WinBUGS code for running the Type II model

```
# use the one-directional conditional distribution for b
model {
for (i in 1:I) {
for (t in 1:T) {
           # Likelihood
                Y[i, t] ~ dpois(mu[i, t])
mu[i, t] <- E[i]*eta[i, t]
log(eta[i, t]) <- alpha  + f[i, t]
 # spline: Z is the tensor product of cubic B-splines
f[i,t] <-  inprod(Z[i, ], b[, t])
    }}

# define number of neighbors
num[1] <- 1
num[2] <- 5
for (t in 3:(T-2)){
num[t] <- 6 }
num[T-1] <- 5
num[T] <- 1
```

```
# second order random walk (RW2) :  undirectional condimental dstribution
 for (k in 1:nknots){
for (t in 1:1){
bmean[k, t] <- (2*b[k, t+1] - b[k, t+2] )}
for (t in 2:2){
bmean[k, t] <- (2*b[k, t-1] +4*b[k, t+1] -  b[k, t+2])/5}
for (t in 3:(T-2)){
bmean[k, t] <- ( - b[k, t-2] + 4*b[k, t-1] +4*b[k, t+1] -  b[k, t+2])/6}
for (t in (T-1):(T-1)){
bmean[k, t] <- (- b[k, t-2] +4*b[k, t-1] +2* b[k, t+1])/5 }
for (t in T:T){
bmean[k, t] <-  ( -b[k, t-2] +2*b[k, t-1])}
}

# calculate the likelihood function of precision parameter tau.b
for (t in 1:T){
tautmp[t] <- num[t] * tau.b
for (k in 1:nknots){
b[k, t] ~ dnorm(bmean[k, t] ,  tautmp[t])
# likelihood of tau
tau.like[k, t]   <- num[t] * b[k, t] * (b[k, t] - bmean[k, t]) }
}

# priors
alpha ~ dnorm(0.0,1.0E-6)
d <- 1+ sum(tau.like[, ]) / 2
r <- 0.005 + nknots*(T-2)/ 2
tau.b  ~ dgamma(r, d)
sigma.b <- 1 / sqrt(tau.b)
}
}
```

Example INLA code for running the Type II model

```
### construct the stacked design matrix: Z matrix is the tensor-product B spline
X.mat <- matrix(NA, nrow=n.area*n.time, ncol=n.basis*n.time)
I.tmp <- diag(1, nrow=n.time, ncol=n.time)
for (i in 1:n.area){
    X.mat[((i-1)*(n.time)+1):(i*n.time),] <- I.tmp %x% matrix(Z[i,], nrow=1)
}
dim(X.mat) # should be (IT)*(KT)

#########--------------- type 2 interaction ---------------#########
# create vectors in the A matrix
intercept <- c(1,  rep(NA, n.basis*n.time))
idx  <-  c(NA, rep(1:n.time, each=n.basis))
group <-  c(NA, rep(1:n.basis, n.time))
data.inla <-  list(y=y, idx=idx, intercept=intercept, group=group)

# fit the model
formula.2 <-  y ~ -1 + intercept +
     f(idx, model = "rw2", constr=F, hyper = list(prec = list(param=c(1, 0.005))),
     replicate=group)
r.2 = inla(formula.2, data=data.inla,family="poisson", control.predictor =
      list(A=cBind(rep(1, nrow(X.mat)), X.mat),
      compute=TRUE, link=1),control.compute=list(dic=T, cpo=TRUE),
```

```
      verbose=F, E=E.lvec)
summary(r.2)

# extract the coefficients
b.hat.2.inla <- r.2$summary.linear.predictor[(n+2):(n+1+m), "mean"]
b.hat.mat.2.inla <- matrix(b.hat.2.inla, nrow=n.basis, ncol=n.time, byrow=F)
head(b.hat.mat.2.inla)
```

5. Simulation

Bubble plots of the observed number of cases in the simulation study can be found in Figure 5.

Table 1. Comparison of the parameter estimates using the Type I–IV models in the simulation study. We report the posterior medians for the intercept α and the variance parameter σ_b^2.

Parameter	Type I	Type II	Type III	Type IV
α	-0.1	-0.09	-0.09	-0.08
σ_b^2	2.6	0.01	4.8	0.15

Table 1 gives parameter estimates from the four interaction models. For all four models, the estimated overall level α is very close to the true value of -0.1. In the table we also report the variance parameter σ_b^2, which is the inverse of the precision parameter τ_b. The estimates of σ_b^2 are very different in the four models, and this is because of the different interpretations of this parameter between the four types and so the parameter is not directly comparable. For example, in the Type II interaction model, σ_b^2 is the variability of the basis coefficients b_{kt} *conditioned* on those at the *neighboring time points* $b_{kt'}, t' \sim t$. In the Type III interaction model, σ_b^2 is the variability of the basis coefficients b_{kt} *conditioned* on those at the *neighboring spatial bases* $b_{k't}, k' \sim k$. Finally, for the Type IV interaction model, σ_b^2 is the variability of the basis coefficients *conditioned* on both the neighboring time points and the neighboring spatial bases.

6. Comparison between MCMC and INLA

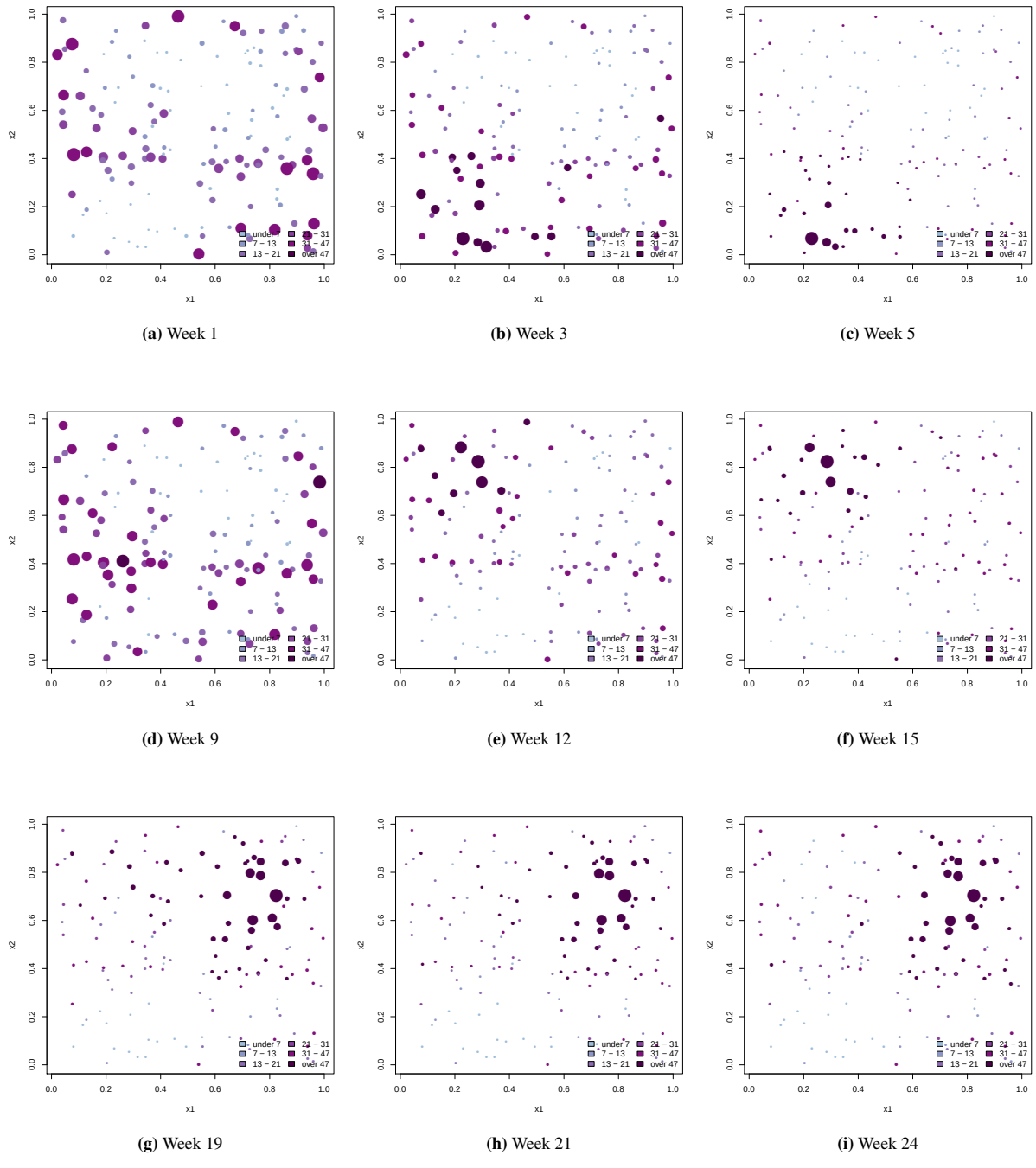
Our proposed models can be fit using Markov Chain Monte Carlo (MCMC), or the newly developed Integrated Nested Laplace Approximation (INLA) technique (12). Here we compare the inferential summaries for each of these two approaches, based on the simulation study described in the main text. To reiterate the model we used for simulation, let Y_{it} indicate the observed count at area i and time t, and E_{it} indicate the expected count calculated using internal standardization (in some cases, we may use the population total N_{it} instead). Then for the relative risk μ_{it} we have

$$Y_{it}|\mu_{it} \sim \text{Poisson}(E_i\mu_{it}) \quad (7)$$

$$\log(\mu_{it}) = \alpha + f_{\text{ST}}(\boldsymbol{x}_i, t), \quad (8)$$

where f_{ST} is the space-time interaction term and constructed using tensor product cubic B-splines. The observed counts Y_{it} are simulated at $I = 150$ randomly selected locations and $T = 24$ discrete time points, with the expected number of cases E_i between 5 and 50 and constant over the study period. The basis coefficients b_{kt} are estimated assuming proposed Type I – IV interaction models.

The MCMC approach is implemented using WinBUGS, and the INLA approach is implemented in R package `INLA`. For each of the interaction models, we run MCMC for $55,000$ iterations after an initial $5,000$ iterations for burn-in. Visual inspection shows good mixing of MCMC chains. Table 2 provides the comparison of parameter estimates between INLA



(a) Week 1 **(b)** Week 3 **(c)** Week 5

(d) Week 9 **(e)** Week 12 **(f)** Week 15

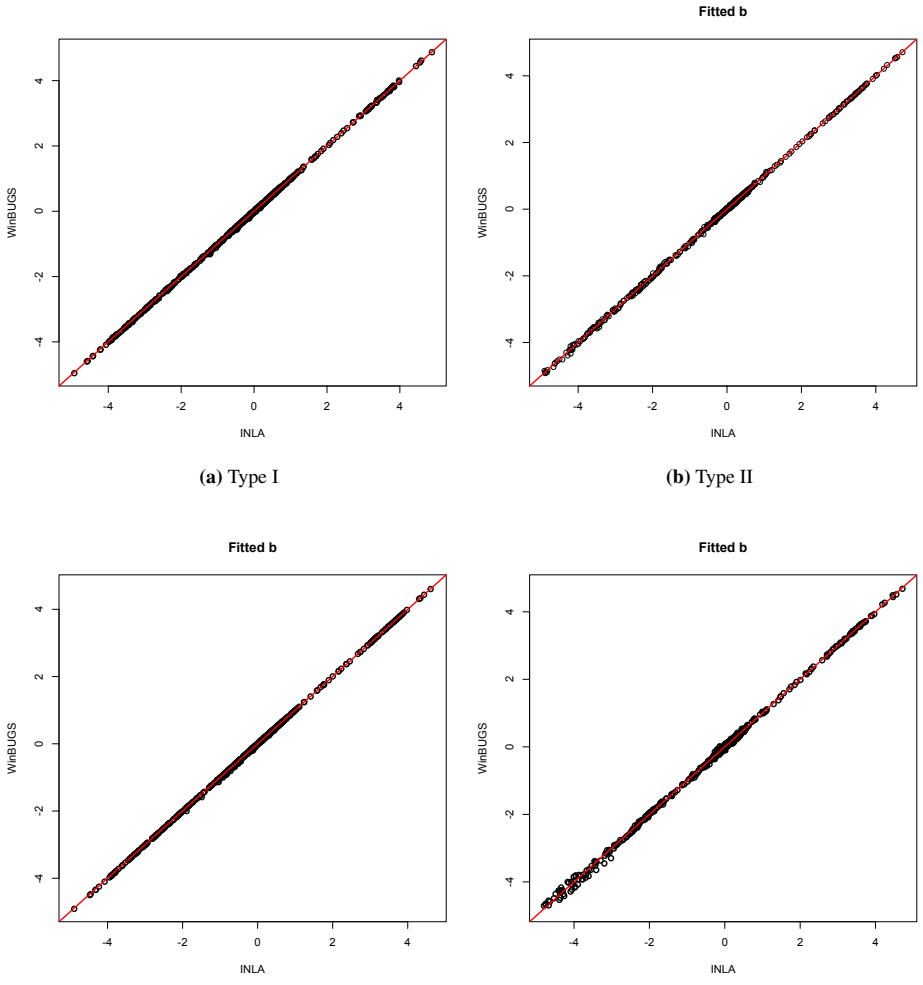
(g) Week 19 **(h)** Week 21 **(i)** Week 24

Figure 5. Bubble plots of the simulated number of cases at selected time points: size of the bubbles are proportional to the number of observed cases.

and MCMC, as well as their respective computation time in seconds. The parameter estimates between INLA and MCMC are very close in terms of both the point estimation and their 95% credible intervals. Figure 6 shows the comparison of estimated basis coefficients b_{kt} using INLA and MCMC, and the estimation again shows agreement. However, shown in Table 2, the computation time of INLA is much shorter than that of MCMC, clearly suggesting the advantage of using INLA here.

Table 2. Comparison of the parameter estimates using the Type I–IV models between MCMC and INLA. We report the posterior medians for the intercept α and the variance parameter σ_b^2, along with their 95% posterior intervals (in parentheses). We also compare the computation time (in seconds) between MCMC and INLA approaches. For each type, we run MCMC for 55,000 iterations with an initial 5,000 iterations as burn-in.

Inference	Parameter	Type I	Type II	Type III	Type IV
INLA	α	-0.099 (-0.13, 0.071)	-0.085 (-0.11, -0.056)	-0.086 (-0.11, -0.058)	-0.084 (-0.11, -0.055)
	σ_b^2	2.57 (2.32, 2.86)	0.11 (0.09, 0.13)	4.88 (4.35, 5.48)	0.16 (0.13, 0.20)
	time (sec)	7.2	15.8	48.1	60.3
MCMC	α	-0.098 (-0.12, 0.071)	-0.084 (-0.11, 0.057)	-0.085 (-0.11, -0.058)	-0.083 (-0.11, -0.054)
	σ_b^2	2.57 (2.32, 2.86)	0.11 (0.09, 0.13)	4.9 (4.46, 5.5)	0.2 (0.14, 0.16)
	time (sec)	9240	9625	10010	10230



(a) Type I **(b)** Type II

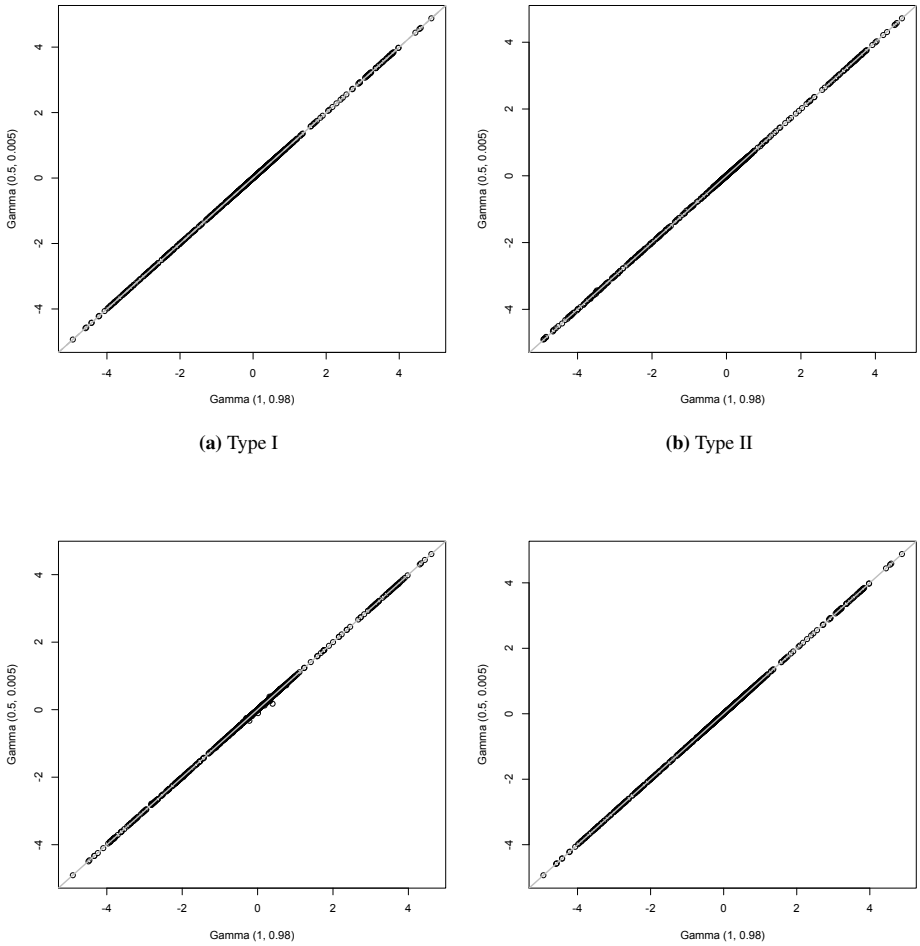
(c) Type III **(d)** Type IV

Figure 6. Comparison of the estimated spline coefficients using four types of interactions in the simulation study, between results obtained with MCMC and INLA.

7. Sensitivity of Precision Parameters Priors

The interaction term in our proposed model has precision parameter τ_b, and is assigned a Gamma distribution prior. To investigate the sensitivity of our model to the prior choices, we compare the estimated spline coefficients $\boldsymbol{b}$ between using Gamma (1, 0.98) and Gamma (0.5, 0.005) as the priors, with the simulated data described in the paper. The results of the

estimation, presented in Figure 7, show little change when using different priors. This exercise confirms that our models are insensitive to the prior choice.



(a) Type I **(b)** Type II

(c) Type III **(d)** Type IV

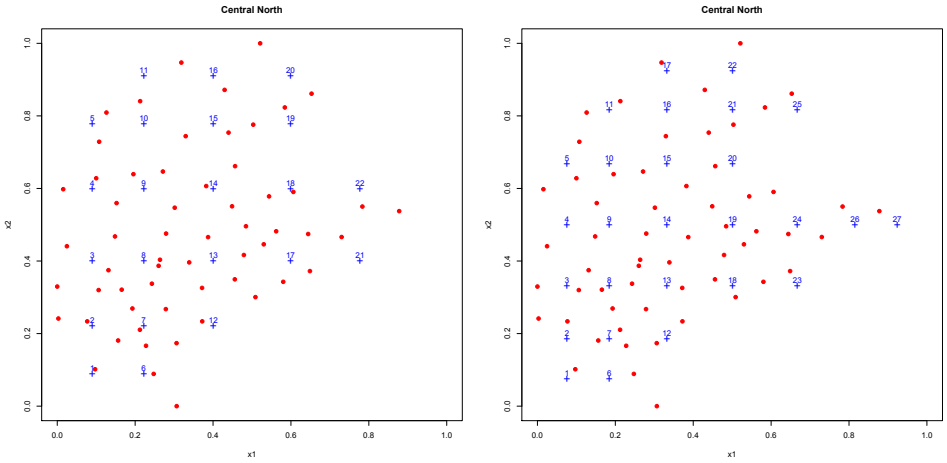
Figure 7. Comparison of the estimated spline coefficients using four types of interactions in the simulation study, with Gamma (1, 0.98) and Gamma (0.5, 0.005) as the priors for the precision parameter.

8. Prior of the Precision Parameter

In the simulation study, the prior of the precision parameter τ_b of the basis coefficients b_{kt} is chosen as Gamma(1, 0.98). The (2.5%, 97.5%) quantile interval of the coefficients b_{kt} from this prior is (-4.8, 4.7) based on 1000 simulations. The resulting relative risk captured by the interaction term at the observed locations has (2.5%, 97.5%) quantile of (0.15, 6.77), which is considered a wide range for the possible values of residual relative risks.

9. Sensitivity of Knots Choice

To compare how sensitive our proposed models to the selection of knots (both in total numbers and locations) we conduct the following sensitivity analysis. In addition to the knots choice presented in the main text, we create a second set of knots, with $k = 27$ total knots (compared to $k = 22$ knots presented in the main text) and different locations, presented in Figure 8. Models with larger number of knots would be expected to perform better, but they also suffer from losing parsimony. Hence we choose the total number of knots $k = 22$ vs $k = 27$ for fair comparison. The parameter estimates (median and 95% interval) from four types of interaction models, using two sets of knots, are shown in Table 3. The estimated coefficient of population density β may be seen as slightly different between these sets of knots. We believe the difference is a result of the high correlation between the fixed effect (i.e. population density) and the random effect, as they share similar spatial structures. This is the circumstance described in detail by (13). When we compare the estimated total counts $\widehat{Y_{it}}$, shown in Figure 9, different knot choices actually provide very similar results. Because the aim of our model is for prediction and not for estimating ecological associations, we see our models insensitive to the knots choice. Between the two sets of knots, we prefer models with $k = 22$ because of our preference to parsimonious Bayesian spatial-temporal models, and so we present results with $k = 22$ in the main text.

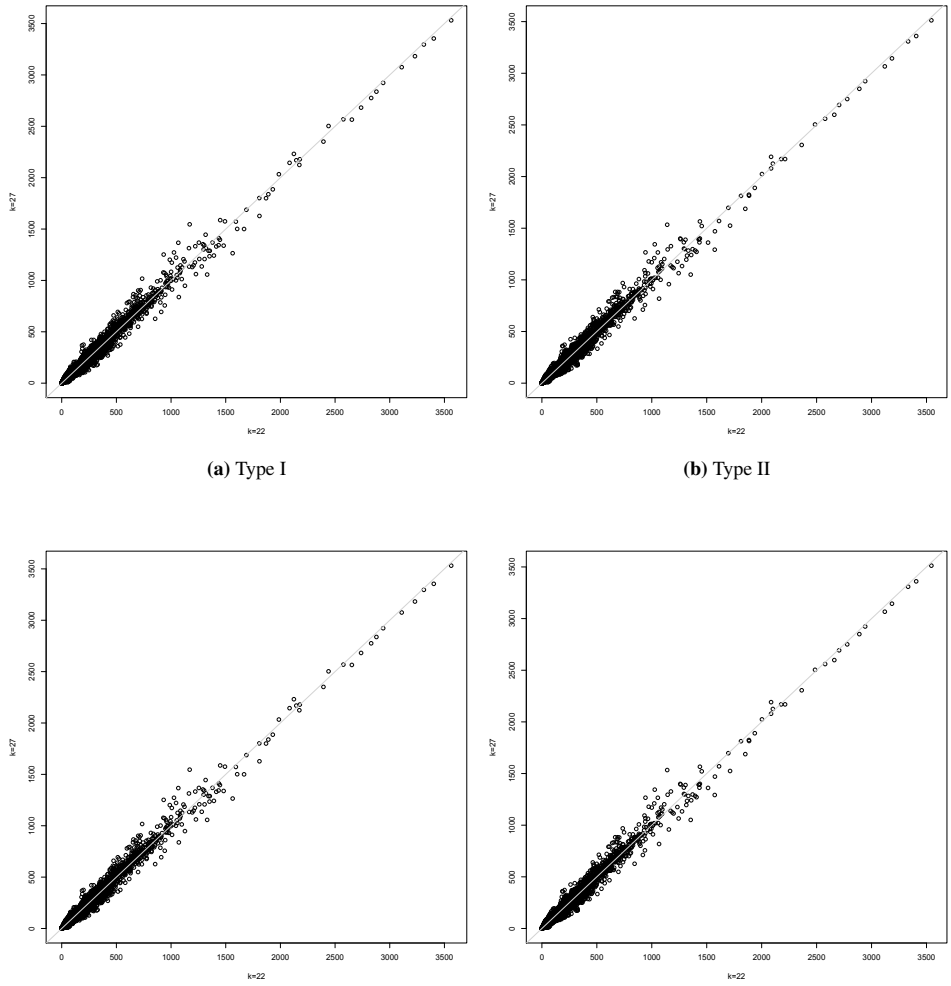


(a) $k = 22$ **(b)** $k = 27$

Figure 8. Two different sets of knots numbers and locations used in the sensitivity analysis of knots choice.

Table 3. Comparison of parameter estimates (median and 95% interval) from four types of interaction models, using weekly central north HFMD data in China 2009–2010.

	Parameter	Type I	Type II	Type III	Type IV
$k = 22$	α	-1.08 (-1.26, -0.88)	-0.97 (-1.45, -0.48)	-0.84 (-1.31, -0.37)	-0.96 (-1.45, -0.46)
	β	0.30 (0.14, 0.45)	0.15 (-0.12, 0.41)	0.33 (0.13, 0.52)	0.16 (-0.11, 0.42)
$k = 27$	α	-1.07 (-1.23, -0.88)	-1.09 (-1.22, -0.96)	-0.9 (-1.5, -0.33)	-0.83 (-1.39, -0.26)
	β	0.25 (0.07, 0.42)	0.27 (0.1, 0.43)	0.30 (0.08, 0.51)	0.28 (-0.06, 0.62)



(a) Type I **(b)** Type II

(c) Type III **(d)** Type IV

Figure 9. Comparison of estimated total counts using two different sets of knots, using weekly central north HFMD data from 2009 to 2010.

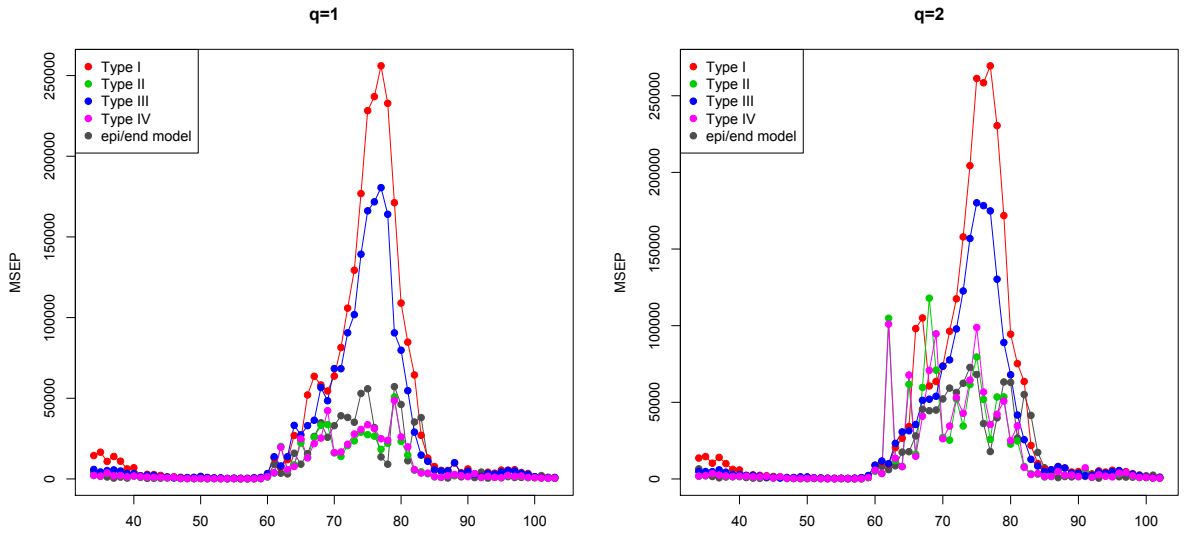
10. Prediction Comparison with China Central North HFMD data

Figures 10 presents MSEP of q-weeks ahead prediction using weekly central north HFMD data from 2009 to 2010. Here we choose q to be 1 and 2. The x-axis represents the week T^* up until which we have the observations. The y-axis represents the MSEP, and are for the Type IIV spline models and the epidemic/endemic (epi/end) model. For each q, the plots suggest that the performance of the Type II and Type IV of our proposed models is comparable to the epi/end model.

11. Results from Epi/End Model with China Central North HFMD data

For comparison we also carry out the analysis of the China HFMD data using the epi/end model of (1),with all weeks between 2009 and 2010. The analysis is implemented within the `R surveillance` package. Specifically, we use the form $y_{it}|\theta_{it} \sim \text{Poisson}(\theta_{it})$ with

$$\theta_{it} = \lambda_i^{\text{AR}} y_{i,t-1} + \lambda_i^{\text{NE}} \sum_{j \neq i} w_{ji} y_{j,t-1} + E_i \lambda_{it}^{\text{EN}} \tag{9}$$



(a) $q = 1$ **(b)** $q = 2$

Figure 10. MSEP of q-weeks ahead prediction using weekly central north HFMD data from 2009 to 2010, using Type I – IV spline models and the epidemic/endemic (epi/end) model.

where $\lambda_i^{\text{AR}} = \alpha_0^{\text{AR}} + b_i^{\text{AR}}$ represents the autoregressive "self-area" component with random effects b_i^{AR}, $\lambda_i^{\text{NE}} = \alpha_0^{\text{NE}} + b_i^{\text{NE}}$ is an autoregressive "neighboring area" component with random effects b_i^{AR}, and

$$\lambda_{it}^{\text{EN}} = \alpha_0^{\text{EN}} + b_i^{\text{EN}} + \beta z_i + \sum_{s=1}^{S} \left(\gamma_s \sin \left[\frac{st}{52} 2\pi \right] + \delta_s \sin \left[\frac{st}{52} 2\pi \right] \right)$$

is an endemic component with the random effects b_i^{EN} and $S = 1$. All the random effects are assumed independently and normally distributed. The results are summarized in Table 4. Clearly the estimated self-area autoregressive $\exp(\alpha^{\text{AR}})$ (i.e., 0.83) is much larger then the neighboring-area autoregressive $\exp(\alpha^{\text{NE}})$ (i.e., 0.05), suggesting the epidemics are driven locally. This is in line with what out models suggest: majority of the variation in the data is from temporal trend rather than the spatial component.

Table 4. Results of the epi/end model, using weekly central north HFMD data in China 2009–2010.

Model	$\exp(\alpha^{\text{AR}})$	$\exp(\alpha^{\text{NE}})$	α_0^{END}	β	σ^{AR}	σ^{NE}	σ^{NE}	l.pen	l.mar
epi/end model	0.83	0.05	-13.42	1.04	0.13	1.38	1.07	-47393.57	-464.34
	(0.015)	(0.0092)	(0.38)	(0.14)					

References

[1] Held L, Hohle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling* 2005; **5**:187–199.

[2] Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics* 1991; **43**:1–59.

[3] Cressie N, Chan N. Spatial modelling of regional variables. *Journal of the American Statistical Association* 1989; **84**:393–401.

[4] Rue H, Held L. *Gaussian Markov Random Fields: Theory and Application*. Chapman and Hall/CRC Press: Boca Raton, 2005.

[5] Clayton D. Generalized linear mixed models. *Markov Chain Monte Carlo in Practice*, Gilks W, Richardson S, Spiegelhalter D (eds.). Chapman and Hall, 1996; 275–301.

[6] Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 2000; **19**:2555–2567.

[7] Green P, Silverman B. *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall/CRC Press: Boca Raton, 1994.

[8] Wood S. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press: Boca Raton, 2006.

[9] Eilers P, Marx B. Flexible smoothing using B-splines and penalized likelihood. *Statistical Science* 1996; **11**:89–121.

[10] Kammann EE, Wand MP. Geoadditive models. *Applied Statistics* 2003; **52**:1–18.

[11] de Boor C. *A Practical Guide to Splines*. Springer: New York, 1978.

[12] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B* 2009; **71**:319–392.

[13] Hodges JS, Reich BJ. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* 2010; **64**(4):325–334.