PLISKA
STUDIA MATHEMATICA

ПЛИСКА
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. **27** (2017), 123–130

PLISKA
STUDIA MATHEMATICA

A BAYESIAN SPATIAL ANALYSIS OF MUMPS DATA IN BULGARIA

Maya Zhelyazkova, Roumyana Yordanova, Maroussia Slavtchova-Bojkova

Bayesian spatial methods have been widely applied in different scientific areas such as epidemiological studies, image processing and many others. In this work we use Bayesian hierarchical model with Gaussian conditionally autoregressive prior to a collection of weekly mumps data from 2007 outbreak in Bulgaria. We generate a disease mapping of the crude standardized incidence ratio across all regional centers. Similar mapping is also produced for the smoothed relative risk. The combination of methods for estimates of the relative risk is a powerful tool to identify high risk regions and may be used to guide local authorities and programs.

1. Introduction

Mumps outbreaks continue to exist in developing world and although rare still occur in vaccinated parts of the world. Spatial autocorrelation analysis is a useful tool to analyze past outbreaks and help health authorities understand the spatial distributions of such outbreaks over time. Several recent papers have analyzed large mumps and measles outbreaks using spatial models. Polgreen et al. [8] applied negative binomial regression to model Iowa mumps epidemic of 2006 and concluded that spring-break college travel was associated with the spread of mumps to other age groups. Porter et al. [9] applied a version of spatial compartmental epidemic model with general latent time distributions to model the same outbreak and came up with similar results showing spring break

2010 *Mathematics Subject Classification:* 62-07.
Key words: Conditional autoregressive models, Disease mapping, spatial correlation.

increased the mixing rate in the population and that the spatial transmission of the disease spreads across multiple conduits. It is worth mentioning that the Iowa outbreak happened in a highly vaccinated population and both the waning of immunization and the social networks of the college student are the main contributing factors. Eccles et al. [2] used Moran's I and local indicators Getis and Ord's G* of spatial association analysis (LISA) to identify clusters of high or low immunization rates. The same methods were also applied to model time changes over space of the immunization rates. Hens et al. [3] used multicohort model for serological information on mumps in a highly vaccinated population of Belgium to quantify the risk of mumps outbreaks in 2012. Lieu et al. [7] studied population of 154,424 children having electronic health records and applied spatial scan statistics to identify five statistically significant clusters of underimmunization and vaccine refusal.

2. Data

There was an outbreak of mumps in Bulgaria in 2007 as discussed in Kojouharova et al. [4]. There were 997 cases registered between January 1st and March 18th of 2007. It affected mostly younger generations between ages 15-19 and 20-24. The main reason for the mumps outbreak was the poor immunisation in these age groups. To deal with the outbreak, the health authorities decided to offer a supplementary mumps immunisation. The mumps data in this study are selected from the first 12 weeks of 2007 collected from 28 different regions of Bulgaria. For every region we have 12 observations which represent the new cases of mumps in consecutive weeks of 2007. The data is provided by the National Center of Infectious and Parasitic Diseases. Our main goal is to reanalyze the mumps data in order to get better understanding of the outbreak.

3. Disease Mapping

Disease Mapping looks for areas with elevated relative risk. There are 28 regional centres and 262 municipalities in Bulgaria. We denote by $O = (O_1, O_2, \ldots, O_n)$ the observed number of cases of mumps in each regional center, $n = 28$ for the first 12 weeks of 2007. We calculate the expected number of cases $E_i, i = 1, \ldots, 28$ by multiplying the number of people in each regional center by the incidence rate of mumps in Bulgaria. The incidence rate is the number of cases of mumps for the first 12 weeks of 2007 divided by the population of Bulgaria. We obtain an incidence ratio $SIR_i = O_i/E_i$ for each region. The incidence ratio is a crude estimate of the relative risk. We then search for regions with elevated risk. If the relative risk is greater than one, then the region is exposed to a higher risk of

the disease. If the relative risk is less than one, the population is comparatively healthy. Sometimes, the risk is elevated by chance due to small E_i because the disease is rare or the population in the area is small. That is why we use Bayesian hierarchical models which decrease the probability of having elevated risk by chance.

4. Bayesian Modeling

In disease mapping spatial autocorrelation is the correlation between close locations on a geographical 2D map which violates the assumptions in classical statistics that assumes independence among observations. If nearby areas share similar patterns we have positive spatial autocorrelation while if they are dissimilar we observe negative spatial autocorrelation. In both cases standard regression models that do not take into account these dependencies will not work since the parameter estimates and results will be unreliable. Since our data is spatially autocorrelated as discussed in Zhelyazkova and Bojkova [11] we use instead hierarchical bayesian models which take into account the prior knowledge about spatial map of Bulgarian regions to achieve a reliable estimate of the mumps risk. Basic ideas in Bayesian modeling is that the parameters within likelihood model are considered random variables with prior distributions. The parameters in the prior distributions can also be stochastic and this establishes a natural parameter hierarchy that leads to hierarchical models. A simple example of hierarchical model that is commonly used in disease mapping is where the data likelihood is Poisson and there is a common relative risk parameter with the single gamma prior distribution. Selecting proper prior distributions is important in building the model.

The main disadvantages of gamma or beta prior for the relative risk in uncorrelated heterogeneity models are that they can not easily model spatially correlated parameters or extend to adjust for covariates. One flexible method to address these drawbacks is proposed by Besag et al. [1]. He models tract count effects by

$$\exp\{x_i'\beta + \nu_i + u_i\},$$

where $x_i'\beta$ is fixed covariate component, ν_i and u_i are uncorrelated and correlated heterogeneity, respectively. Each of the components in the model has different prior distribution. The correlated heterogeneity often follows conditionally autoregressive (CAR) prior distribution.

5. Convolution model

We use one of the Bayesian hierarchical models, so-called convolution model, described in Lawson [5].

We begin with the following notation. Let W be the neighbourhood matrix, where $w_{ij} = 1$, if i and j are neighbours or if the areas i and j share a common border, otherwise $w_{ij} = 0$. We denote that i and j are neighbours by $i \sim j$.

We fit the model $\theta_i = \exp(\beta_0 + \nu_i + u_i)$, where θ_i is the Standardized Incidence Ratio (SIR), which can be interpreted as a relative risk, β_0 is the model intercept with a non-informative prior, ν_i are the structured random effects that assume a normal distribution with variance σ_ν^2 and u_i are the random effects that capture the spatial autocorrelation between regions in a Gaussian CAR model with variance σ_u^2. More precisely the model is

$$O_i|E_i, \theta_i \sim Poisson(E_i\theta_i), i = 1, \ldots, n$$

$$\ln(\theta_i) = \beta_0 + \nu_i + u_i$$

$$\nu_i|\sigma \sim N(0, \sigma_\nu^2)$$

$$u_i|u_j, j \neq i, W, \sigma_i^2 \sim N\left(\frac{1}{n_i}\sum_{i\sim j} u_j, \frac{\sigma_u^2}{n_i}\right)$$

This model accounts for overdispersion and this is an advantage over the pure Poisson model. The Poisson model induces $Var(O_i) = E(O_i)$, but in most of these studies $Var(O_i) > E(O_i)$ or we have overdispersion.

6. Results

We apply the convolution model to the mumps data from the 2007 mumps outbreak in Bulgaria. We use the cumulative data from 28 regional centres. There are 12 consecutive observations per region each corresponding to weekly reports of mumps in 2007.

The results were obtained using CARBayes (Duncan Lee, [6]) package of software "R". Summary of the estimated parameters from the model is presented in Table 1. Inference for the model is based on 80,000 MCMC simulations obtained by running the chain for 100,000 samples where 20,000 were discarded as burn-in period.

The high density posterior intervals does not contain zero and the random effect variance due to spatial autocorrelation σ_u is much larger than the variance of structure random effects σ_v. We also calculated the Deviance Information

Table 1: Model estimates and 95% high density posterior intervals (HDPI)

	Median	2.5 %	97.5 %
intercept β_0	-0.7892	-0.8620	-0.7214
σ_u^2	3.4644	2.4786	4.6446
σ_ν^2	0.1773	0.0371	0.4085

Criterion (DIC) and the effective number of parameters (pD), see Spriegelhalter et al. [10]. The DIC=1566.6615 and pD=227.2521. pD is close to the number of the true parameters 262, assuming independence of the ratios.

In order to check whether the convolution model is appropriate we apply the Moran's I test to the residuals of the convolution model as discussed in Lawson [5]. If the p-value of the Moran's I test is high it means that there is no spatial autocorrelation in the residuals and consequently the convolution model works well. In our case p-value = 0.9338 which is interpreted as a good model fit.

The maps in Figures 1 and 2 give a visual indication of the degree of risk of morbidity in different areas. The incidence ratio $SIR_i = O_i/E_i$ for each region in Figure 1 is calculated from the observed data where the expected number of cases E_i, $i = 1, \ldots, 28$ is equal to the number of people in the regional centre multiplied by the incidence rate of mumps in Bulgaria. The model estimate of the relative risk θ_i is plotted in Figure 2.

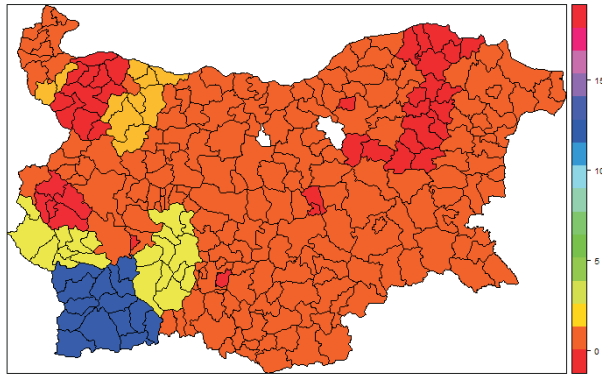

Figure 1: SIR ratio on the map of Bulgaria across all 262 municipalities. It is plotted applying sppplot function as in Lee [6].

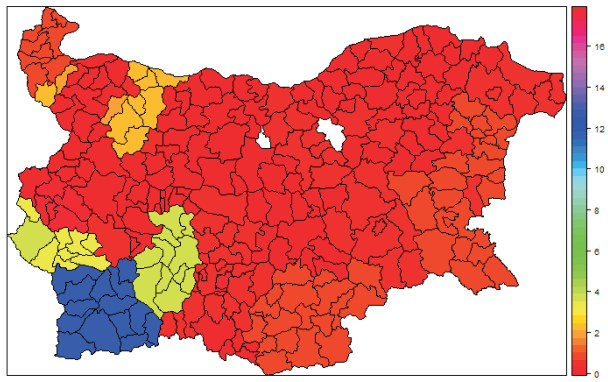

Figure 2: Model estimate of the relative risk from the convolution model on the map of Bulgaria. The magnitudes of the relative risk are mapped into the colour bar.

The relative risk in Figure 2 is smoother in comparison to the one in Figure 1. The reason is that the convolution model makes a smooth transition between the risk of morbidity in different areas because is takes into account the influence of their neighbours. Along with the analytical part the model gives a visual representation of the risk in each municipality of the country. In this way we get a more complete and accurate picture of the morbidity risk.

Our analysis shows that the highest risk of mumps is in the southwest part of Bulgaria. These are mountain regions at the border of Bulgaria. There is also a pomak population with higher percentage of younger children which facilitate the mumps spread. We have found in preliminary studies that mumps outbreaks are due to poor immunization culture of minorities. However, this hypothesis is not confirmed here as we found small non significant negative correlation between the number of cases and the percent of gypsy population in Bulgaria. The results might be different if there was not a suspension of mumps immunization for the whole population of Bulgaria for the time period 1982–1986 as stated in Kojouharova et al. [4].

7. Discussion

The convolution model takes into account the influence of neighbouring areas on the outcome of the selected area, which is closer to the real picture in studying the spread of infectious disease. This is evident in the analysis of incidence in different

areas of the country. The SIR mapping is a crude approximation and sometimes is hard to interpret. That is why we consider a smooth version which is disease mapping of the relative risk as a model estimate from the convolution model. On the other hand the smooth mapping in some cases can produce oversmoothing and it will not reveal all the features of the relative risk.

All the results in this paper are scale dependent. We consider 28 regional centres, but we may as well fit the model on 262 municipalities for which we expect better local risk estimation.

In order to reduce the risk of future mumps outbreaks the health authorities need to work on the prevention and the immunisation culture of the nation. As a future research we could include more covariates in the convolution model, so that we have better overview of the mumps outbreak. Another direction of future investigation would be to fit negative binomial model to the mumps data and to compare the results with the convolution model.

8. Acknowledgements

The shapefiles used in this paper were provided by Dr. Anton Popov from the Department of Geology and Geography at Sofia University. This work is partially supported by the National Fund for Scientific Research at the Sofia University (grant 80-10-146/21.04.2017).

REFERENCES

[1] J. Besag, Y. York, A. Mollié. Bayesian image restoration with two applications in spatial statistics. *Ann. Inst. Stat. Math.*, **43**, No 1 (1991), 1–59.

[2] K. M. Eccles, S. Bertazzon. Application of geographic information system in public health: A geospatial approach to analyzing MMR immunzation uptake in Alberta. *Can. J. Public Health*, **106**, No 6 (2015), 355–361.

[3] N. Hens, S. Abrams, E. Santermans, H. Theeten, N. Goeyvaerts, T. Lernout, E. Leuridan, K. Van Kerckhove, H. Goossens, P. Van Damme, P. Beutels. Assesing the risk of measles resurgence in a highly vaccinated population: Belgian anno 2013. *Euro Surveill.*, **20**, No 1, (2015), 20998.

[4] M. Kojouharova, A. Kurchatova, L. Marinova, T. Georgieva. Mumps outbreak in Bulgaria, 2007 a preliminary report. *Euro Surveill.*, **12** (2007), 3126.

[5] A. B. Lawson. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Boca Raton, Taylor and Francis Group, 2008.

[6] D. Lee. CARBayes: An R package for Bayesian spatial modelling with Conditional Autoregressive priors. *Journal of statistical software*, **55**, No 13 (2013).

[7] T. A. Lieu, G. T. Ray, N. P. Klein, C. Chung, M. Kulldorff. Geographic clusters in underimmunization and vaccine refusal. *Pediatrics*, **135**, No 2 (2015), 280–289.

[8] P. M. Polgreen, L. C. Bohnett,M. A. Pentella,J. E. Cavanaugh. A spatial analysis of the spread of mumps: the importance of college students and their spring-break-associated travel. *Epidemiol. Infect.*, **138**, No 3 (2010), 434–441.

[9] A. T. Porter, J. J. Oleson. A spatial epidemic model for disease spread over a heterogeneous spatial support.*Stat. Med.*, **35**, No 5 (2016), 721–733.

[10] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van der Linde. Bayesian measures of Model Complexity and Fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**, No 4 (2002), 583–639.

[11] M. Zhelyazkova, M. Slavtchova-Bojkova. Spatial Autocorrelation in Mumps Data in Bulgaria. BGSIAM'14 Proceedings, 2015, 93–101.

Maya Zhelyazkova
e-mail: `zhelyazkova@fmi.uni-sofia.bg`
Maroussia Slavtchova-Bojkova
e- mail: `bojkova@fmi.uni-sofia.bg`
Department of Probability, Operational Research and Statistics
Faculty of Mathematics and Informatics
Sofia University
5, J. Bourchier Blvd
1164 Sofia, Bulgaria

Roumyana Yordanova
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Acad. Georgi Bonchev Str., Bl. 8
1113 Sofia, Bulgaria
e-mail: `ryordanova@math.bas.bg`