© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Spatio-temporal Models Using R-INLA with Generalized Extreme Value Distribution in Hierarchical Bayes Regression

Ro'fah Nur Rachmawati[1,2], Anik Djuraidah*[1], Anwar Fitrianto[1], I Made Sumertajaya[1]
*[1]Statistics Department, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia
[2]Statistics Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

ABSTRACT

Spatio-temporal data containing information about space, time and its interaction allows researchers to describe potential geographical pattern. Bayesian method commonly used in describing spatio-temporal data is simulated based Markov Chain Monte Carlo (MCMC). However, MCMC may be extremely slow in the posterior inference simulation process and it becomes computationally unfeasible if the specified models are complex and designed hierarchically. The Integrated Nested Laplace Approximation (INLA) in R-INLA package is approximated based method and becomes a viable alternative to fundamental limitation of the expensive MCMC computation. This paper aims to model data using Bayes spatial and spatio-temporal models divided in parametric and non-parametric temporal trend specifications. We use the number of poor people as the response that fit to generalized extreme value distribution and investigate the geographical patterns among regions by adding the socioeconomics information data set in Bayes spatial model. In Bayes spatio-temporal models we conclude classical parametric temporal trend as the best model that can describe space-time interaction based on the smallest deviance criteria. All the estimation processes are performed efficiently with R-INLA resulting fast, accurate and guarantee of convergence posterior inferences compared to MCMC's convergence issues.

Keywords: Areal Data, Gaussian Markov Random Field, Laplace Approximation, Random Walk

I. INTRODUCTION

The advances in computational tools allows researcher to collect real-time data from satellites, GPS, etc. This means, a great opportunity of geo-reference data that contain information about space or also time, increase the possibility to analyze highly multivariate data, with many important response variables and predictors. Furthermore, geographically reference data that presented as maps in longitudinal data or other time series structure, enlarge the possibilities of the researchers to capture the temporal correlated structure from time to time (Banarjee et al., 2015). As an example, if we consider to the epidemiological investigation, we might interest to analyze the incidence of certain disease such as lung, breast or cervical cancer rates in given counties, with information available such as smoking, food behavior, mammography and other important information at the same level. It is very important to consider the potential geographical pattern of the disease that is areas are close to each other are more likely share the same geographical characteristics which are related to the disease, thus to have similar incidence (Blangiardo and Cameletti, 2015). However, investigating only the spatial pattern of disease does not allow us to say anything about their temporal variation which could be equally or even more important and interesting. Spatio-temporal modeling become one of space-time analysis method that widely used to describe geo-referenced data in diverse areas from climatology (Carvalho et al., 2016; Aravena and Luckman, 2009) to environmental health (Yanosky et al., 2014; Rui Ye et al., 2014; Abellan et al., 2008; Zongwei Ma et al., 2016), from ecology (Djeckman et al., 2009) to epidemiology (Korennoy et al., 2014; Nazia et al., 2017).

Bayesian inferences have developed greatly and widely established in many scientific fields such as social science (Kaplan, 2014; Jackman, 2009) to health economic assessment (Baio et al., 2017), and from exposure science to epidemiology (Pirani et al., 2014). In Bayesian perspective, there is no fundamental difference between unobservable data and unknown parameters (or also hyperparameters) that we also consider as random quantities. Therefore, usually but not necessarily, the objective of Bayesian inference is using the combination of data and current information of parameters described with prior distribution to derive its posterior distribution. Determination of prior distribution can be based on informative prior such as previous studies or expert opinion and from non-informative prior when there is no available information about the characteristics of parameters. In non-informative cases, conjugate prior often used as an option to prior distribution selection to the set of parameters that characterized with large variability. There are several excellences in Bayesian inferences, within Bayesian approaches it is easy to define hierarchical structure on the data and/or parameters, which presents the additional benefit on small or missing sample prediction. Moreover, Bayesian inferences generally also have good frequentist properties, because frequentist method is a special case of Bayesian inference with certain prior distribution.

Bayesian method deals with spatial and spatio-temporal modeling, with the development of simulative method of Markov Chain Monte Carlo (MCMC) (Casella and George, 1992; Gilks et al., 1996). Spatial and spatio-temporal models are specified in Bayesian framework by simply extending the concept of hierarchical structure, allowing to share the strength of geo-reference characteristics based on the neighborhood for area level data or on the distance for point level data, and allowing temporal effect for data with space-time observations. The main contribution to spatial and spatio-temporal statistics is Besag et al. (1991), who developed the Besag–York–Mollié (BYM) method which is commonly used for disease mapping models for area level data, while Banerjee et al. (2004), Diggle and Ribeiro (2007) and Cressie and Wikle (2011) have developed and concentrated on Bayesian geostatistical (point level) models. The appearance of MCMC has greatly increased the possibility to develop complex model with large data set without the need of forcing the model to the simplified structure. MCMC method are powerful computational tool for Bayesian inference and applied in WinBUGS (Ntzoufras, 2009) as part of BUGS (Bayesian inference Using Gibs Sampler) software for the Bayesian analysis. In WinBUGS user only needs to specify a proto-type model, provide the data and the initial values, then the software will automatically select the sampling methods to generate values from the posterior distributions of the specified model. MCMC theory states that the distribution of simulated values converges to the targeted distribution (i.e the posterior distribution) when the iteration number goes to infinity, while it is not feasible to run Markov Chain infinitely. In other words, we do not know how fast the rate of generated samples convergence to the posterior distribution. This is equivalent to asserting that there is no guarantee that Markov Chain will converge, moreover when models are complex (e.g. big n problem such as spatio-temporal models) and designed with a hierarchical fashion, then MCMC maybe computationally unfeasible.

A current development and viable alternative to MCMC methods able to reduce the computational costs of Bayesian inference is INLA (Integrated Nested Laplace Approximation) algorithm (Rue et al. 2009). The INLA is a deterministic analytical approach rather than simulation based as in MCMC.

The INLA designed with latent Gaussian model and provides accurate results produced in much shorter time than MCMC. R-INLA is the R package to implement approximate Bayesian inference using the INLA approach (Martino and Rue, 2010) with has flexible model and very wide class ranging from (generalized) linear mixed to spatial and spatio-temporal models (Blangiardo et al., 2013). For this reason, the INLA is successfully used in great variety of applications such as Ruiz-Cárdenas et al. (2012) that fit the dynamic model using INLA; Martino et al. (2011) applicate INLA to the financial field for estimating the stochastic volatility models; and Simpson et el. (2016) using INLA for efficient computational to fit the log-Gaussian Cox process. In spatial and spatio-temporal analysis, there are several research publications of INLA, such as Blangiardo et al. (2013) using INLA to model the number of suicides into spatial model which assume Poisson distribution in the likelihood of the area level data; Cosandy-Godin et al. (2014) applies spatio-temporal models using INLA combined with Stochastic Partial Differential Equation (SPDE) to fisheries bycatch in the Canadian Arctic, which assume Poisson and Negative Binomial in the likelihood for point level data; Cameletti et al. (2012) using INLA and SPDE to model hierarchical spatio-temporal model for particulate matter (PM) concentration in the North-Italian region Piemonte during winter season October 2005 - March 2006, which assume Gaussian in the likelihood of point level data; and Papiola et al. (2013) using INLA to spatio-temporal analysis in Stomach cancer incidence in Southern Portugal year 1998-2006, which assume Poisson distribution in likelihood of area level data.

This far, the application of INLA in Bayes spatio-temporal analysis are limited and commonly used to assume Gaussian and Poisson likelihood in area or point level data, while the extreme value distribution which has many applications in geo-reference extreme weather data such as rainfall (Aravenaa and Luckman, 2009), wind speeds (Mahmoudian and Mohammadzadeh, 2014) and precipitation events (Ghosh and Mallick, 2011) still use simulation based on MCMC in its estimation method. In financial fields the extreme values distribution can be represented in risk, large fluctuations in exchange rate and market crashes. In socioeconomics, extreme value distribution can be represented as salary income in a country (see Pindado et al. 2017), the number of poor people in given regions divided with rural or urban status and the consumer behavior of people life style who live in city and villages regions. Because the limitation or even no specific literatures describe the generalized extreme value with INLA method, this paper aims to: (1) give the illustration of INLA's estimation using the latent Gaussian model and Gaussian Markov Random Field property as the foundation of the excellence of INLA in computational benefit aspect rather than MCMC. The INLA's illustration is given in section 2.4; (2) find the best model to the number of poor people data set in East Java province, Indonesia that fit generalized extreme value distribution, using Bayes spatial and spatio-temporal models using R-INLA package. The R code, model analysis and results detail explained in section 3. While the theoretical background of Bayes spatial, spatio-temporal models and generalized extreme value distribution, will be explained briefly in section 2.1-2.3; and discussion in section 4 is the summary of the whole models used in this paper, with some review and explanation of the limitation of this paper and at the same time become the opportunity for further research analysis.

II. THEORETICAL BACKGROUND AND METHODS

2.1 Bayes spatial model

Spatial data are defined as stochastic process realizations indexed by space

$$Y(s) \equiv \{y(s), s \in D\},$$

where D is a (fixed) subset of d-dimensional of real number, $\mathbb{R}^d$. In area level data, $y(s)$ is a random aggregate value over an areal unit s with well-defined boundaries in D, which is define a countable collection of d-dimensional spatial units. Spatial dependency is taken from the neighborhood structure, by simplifying the notation, so that $(s_1, s_2, \dots, s_n)$ becomes $(1,2,\dots,n)$, furthermore if given the area i, $\mathcal{N}(i)$ is the set of first order (or second order) neighbor(s) of area i.

An important property in spatial analysis is that specific area shares the strength geostatistical characteristics with its neighbors, thus the area which close to each other tend to have the same characteristics than areas that are far apart. So that the parameter θ_i for the ith area is independent of all other parameters, given the set of its first order neighbors $\mathcal{N}(i)$, in other words, for any pair of elements (i,j) in $\boldsymbol{\theta}$

$$\theta_i \perp \theta_j | \boldsymbol{\theta}_{-ij} \Leftrightarrow Q_{ij} = 0,$$

where $\boldsymbol{\theta}_{-ij}$ indicates all elements in $\boldsymbol{\theta}$ but the i,jth and $\boldsymbol{Q}$ is the precision matrix of $\boldsymbol{\theta}$ is sparse, thus $Q_{ij} \neq 0$ only if $j \in i, \mathcal{N}(i)$, which produces great computational benefits. This specification called Gaussian Markov Random Field (GMRF) (Rue and Held, 2005).

Besag–York–Molliè (BYM) model presented by Besag et al. (1991) specifies the linear predictor as

$$\eta_i = b_0 + \vartheta_i + v_i,$$

where b_0 is the intercept, represents the average outcome in the entire study of region, while ϑ_i and v_i is spatially structured risk effect and spatially unstructured risk effect for specific area i respectively. In addition, Besag, et al. (1974) specifies this model using exchangeable prior for v_i, while in BYM specification the spatially unstructured component is set to be the Normal prior distribution, i.e. $v_i \sim \text{Normal}(0, \sigma_v^2)$, with σ_v^2 is variance parameter that presents the amount of spatially unstructured random effect. We will focus on intrinsic conditional autoregressive one as implemented in R-INLA, so that spatially structured component of (1) stated as

$$\vartheta_i | \boldsymbol{\vartheta}_{-i} \sim \text{Normal}\left(\frac{1}{\#\mathcal{N}(i)} \sum_{j=1}^{n} c_{ij} \vartheta_j, s_i^2 \right),$$

where c_{ij} represents the neighboring criteria with for $i \neq j$, $c_{ij} = 1$ if area i and j are neighbors, and 0 otherwise, and $s_i^2 = \sigma_\vartheta^2 / \#\mathcal{N}(i)$ is the variance for (3) area i depends on its numbers of neighbors, $\#\mathcal{N}(i)$. The variance in BYM specification as in equation (2) implicitly state that if an area has many neighbors then the variance will be smaller, while the variance parameter σ_ϑ^2 controls the amount of variation between the spatially structured random effects (Blangiardo and Cameletti, 2015). So that, the parameters and hyperparameters to be estimated are $\boldsymbol{\theta} = \{b_0, \boldsymbol{\vartheta}, \boldsymbol{v}\}$ and $\boldsymbol{\psi} = \{\tau_\vartheta, \tau_v\}$ respectively, with $\tau = 1/\sigma^2$ defines its precision or the invers of variance parameters.

2.2. *Bayes spatio-temporal models*

As stated in Herrmann et al. (2015), there is a significant trend numbers of cancers in Switzerland which almost six times reduction in the last 40 years, and in Cosandy-Godin et al. (2014) stated that there are certain types of fish will be in a certain place at a certain time as well. Example of such (1) research suggests that there is a real relationship between a (1) phenomenon over time. However, section 2.1 only allows us to evaluate spatial risk pattern in single specified year, we are not able to say anything about the temporal trend of risk for a series of observation years. In spatio-temporal models we can now investigate a temporal trend, and as in spatial process data are now defined by a process indexed (2) by space and time

$$Y(s,t) \equiv \{y(s,t), (s,t) \in D \subset \mathbb{R}^2 \times \mathbb{R}\}$$

observed at n spatial locations or areas and at T time points.

For poin level data, as described in Gneiting et al. (2006) to overcome the computational complexity, the spatio-temporal models simply assume (2) separability in space-time covariance function, so that the covariance can be decomposed into the sum (or product) of a purely spatial and purely temporal term, i.e. $\text{Cov}(y_{ij}, y_{uv}) = C_1(\Delta_{ij})C_2(\Delta_{uv})$, where $C_1(\Delta_{ij})$ and $C_2(\Delta_{uv})$ are covariance function for purely spatial in between location i and j and covariance function for purely temporal between time u and v, respectively. In area level data, GMRF can be applied to include a precision matrix defined also in terms of time and assuming a neighboring structure. If space-time interaction is included in the model (Clayton, 1996; Knorr-Held, 2000), its precision can be obtained through the Kronecker product.

2.2.1 Classical parametric trend

BYM specification as in equation (1) is extended to allow temporal term, as in Bernardinelli et al. (1995) presents a parametric trend for the temporal component which assumes that the linear predictor can be written as

$$\eta_{it} = b_0 + \vartheta_i + v_i + (\beta + \delta_i)t,$$

where the main spatial effect is BYM specification as stated in section 2.1, β is the global temporal effect for all areas and δ_i as differential trend specified for area i for year t. δ_i is set to be exchangeable prior, while in this paper we assume as $\text{Normal}(0, \sigma_\delta^2)$ with the precision is defined with $\tau_\delta = 1/\sigma_\delta^2$. The parameters and hyperparameters to be estimated in this model are $\boldsymbol{\theta} = \{b_0, \boldsymbol{\vartheta}, \boldsymbol{v}, \beta, \boldsymbol{\delta}\}$ and $\boldsymbol{\psi} = \{\tau_\vartheta, \tau_v, \tau_\delta\}$ respectively, with $\tau = 1/\sigma^2$ still defines its precision.

2.2.2 Dynamic nonparametric trend

Implicitly in model (3) we assume linearity constraint in its differential trend, δ_i. In nonparametric model, allows us to model the temporal trend using dynamic formulation as introduced in Knorr-Held (2000)

$$\eta_{it} = b_0 + \vartheta_i + v_i + \rho_t + \varphi_t.$$

This model replaces the global temporal effect and differential trend with two others term, i.e. the structured temporal trend ρ_t and the unstructured temporal trend φ_t, while b_0, ϑ_i and v_i is BYM specification as in previous models. The structured temporal trend modeled dynamically using random walk of order one (or two), defined as

$$\rho_t|\rho_{-t} \sim \begin{cases} \text{Normal}(\rho_{t+1}, \tau_\rho), & \text{for } t = 1 \\ \text{Normal}\left(\dfrac{\rho_{t-1} + \rho_{t+1}}{2}, \dfrac{\tau_\rho}{2}\right), & \text{for } t = 2, \dots, T-1 \\ \text{Normal}(\rho_{t-1}, \tau_v), & \text{for } t = T \end{cases} \quad (5)$$

while $\tau_\rho = 1/\sigma_\rho^2$ is the precision and φ_t is assumed by means of Gaussian exchangeable prior, i.e. $\varphi_t \sim \text{Normal}(0, 1/\tau_\varphi)$. The parameters and hyperparameters to be estimated in this model are $\boldsymbol{\theta} = \{b_0, \boldsymbol{\vartheta}, \boldsymbol{v}, \boldsymbol{\rho}, \boldsymbol{\varphi}\}$ and $\boldsymbol{\psi} = \{\tau_\vartheta, \tau_v, \tau_\delta, \tau_\varphi\}$ respectively.

2.2.3 Space-time interaction nonparametric trend

Expanding model (4) and allowing interaction between space and time is easily formulated using the following specification

$$\eta_{it} = b_0 + \vartheta_i + v_i + \rho_t + \varphi_t + \gamma_{it}. \quad (6)$$

The first five elements in (6) using the same specification as model (4), with space-time interaction effect γ_{it} assumed follows a Gaussian distribution, i.e. $\gamma_{it} \sim \text{Normal}(0, 1/\tau_\gamma)$, with the structure matrix $\boldsymbol{S}_\gamma$, identify the type of spatial and temporal dependence between the elements of $\boldsymbol{\gamma}$. Clayton (1996) using $\boldsymbol{S}_\gamma$ as the Kronecker product of structure matrixes interaction, so if we assume there is an interaction between the spatially unstructured effect and unstructured temporal trend, then we can write the structure matrix as $\boldsymbol{S}_\gamma = \boldsymbol{S}_v \otimes \boldsymbol{S}_\varphi = \boldsymbol{I} \otimes \boldsymbol{I} = \boldsymbol{I}$, because both $\boldsymbol{v}$ and $\boldsymbol{\varphi}$ have no spatial and temporal structure.

2.3 Generalized extreme value (GEV) distribution

Extreme value theory plays an important role in socioeconomics phenomena. Statistical model with extreme value data enable us to make forecasts about the occurrences of such events with fairly high accuracy. Such forecasts can help to minimize the loss of human and economic assets caused by such extreme phenomenon. This fact motivates researchers to develop predictive models for extreme events. The GEV known as Fisher-Tippet distribution is a family of continuous distribution developed with extreme value theory which combined type I, type II and type III extreme value distribution also known as Gumbel, Fréchet and Weibull families respectively. Unlike the normal distribution that arises from the use of the central limit theorem on sample averages, the extreme value distribution arises from the limit theorem of Fisher and Tippet (1928) on extreme values or maxima in sample data. The GEV distribution is defined through the cumulative distribution function

$$F(y;\eta,\tau,\xi) = \exp\left(-\left[1+\xi\sqrt{\tau s}(y-\eta)\right]^{-1/\xi}\right)$$

for

$$1+\xi\sqrt{\tau s}(y-\eta) > 0$$

and for continuously response y where η is the linear predictor, τ is the "precision" defined with the inverse of its variance parameter i.e. $1/\sigma^2$ and s is a fixed scaling, $s > 0$. In Bayesian perspective, GEV has two hyperparameters, i.e. the "precision" is represented as $\theta_1 = \log\tau$, and the prior is defined on θ_1. The shape parameter ξ is represented as $\theta_2 = \xi$ and the prior is defined on θ_2.

2.4 INLA Estimation Method

Purposed by Rue, et al. (2009), INLA algorithm is an analytical Bayesian inference using Laplace approximation method which designed with latent Gaussian models and provides accurate results and shorter computing time compared to MCMC. Suppose that $\boldsymbol{y}$ is the vector of observation with specified prior distribution of parameter θ, Bayesian inference is concentrated around the posterior of its parameter i.e.

$$p(\theta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\theta)p(\theta).$$

The entire part of this section is a brief illustration of INLA estimation. For example, we suppose the specified prior density function is Gamma distribution with

$$g(\theta) = \frac{\alpha}{\Gamma(\alpha)}\exp(-\beta\theta)\,\theta^{\alpha-1}, \qquad \theta,\alpha,\beta > 0.$$

Therefore, we interested in computing the integral

$$\int g(\theta)d\theta = \int \exp(\log g(\theta))d\theta$$
$$= \int \exp\big((\alpha-1)\log\theta - \beta\theta + C\big)d\theta, \tag{7}$$

while C is constant. Present of $\log g(\theta)$ can be approximated using means of Taylor series expansion, which evaluated in $\theta = \theta_0$:

$$\log g(\theta) \approx \log g(\theta_0) + (\theta-\theta_0)\frac{\partial \log g(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\bigg|_{\theta=\theta_0}$$
$$= (\alpha-1)\log\theta_0 - \beta\theta_0 + (\theta-\theta_0)\frac{\partial \log g(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\bigg|_{\theta=\theta_0} + C$$

If θ_0 is set to be its mode, so that $\theta^* = \operatorname{argmax}_\theta \log g(\theta)$, then $\frac{\partial \log g(\theta)}{\partial\theta}\Big|_{\theta=\theta^*} = 0$ results that $\theta^* = \frac{\alpha-1}{\theta} - \beta$, and the approximation becomes

$$\log g(\theta) \approx (\alpha-1)\log\theta^* - \beta\,\theta^* + \frac{(\theta-\theta^*)^2}{2}\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\bigg|_{\theta=\theta^*} + C.$$

The integral of interest in (7) can be approximated as

$$\int g(\theta)d\theta \approx \int \exp\left((\alpha-1)\log\theta^* - \beta\,\theta^* + \frac{(\theta-\theta^*)^2}{2}\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\bigg|_{\theta=\theta^*} + C\right)d\theta,$$
$$= \exp\big((\alpha-1)\log\theta^* - \beta\,\theta^* + C\big)\int \exp\left(\frac{(\theta-\theta^*)^2}{2}\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\bigg|_{\theta=\theta^*}\right)d\theta$$

where the above integrand can be associated with the density of Normal distribution. By setting $\sigma^{2*} = {-1}\Big/\frac{\partial^2 \log g(\theta)}{\partial\theta^2}\Big|_{\theta=\theta^*}$ results that $\sigma^{2*} = \frac{\alpha-1}{\beta^2}$, we obtain

$$\int g(\theta)d\theta \approx \exp\big((\alpha-1)\log\theta^* - \beta\,\theta^* + C\big)\int \exp\left(-\frac{(\theta-\theta^*)^2}{2\sigma^{2*}}\right)d\theta,$$

which associated with the kernel of Normal distribution with mean equal to θ^* and variance σ^{2*}. If the integral is evaluated in the interval (a, b), then the Laplace approximation is

$$\int_a^b g(\theta)d\theta \approx \exp\big((\alpha - 1)\log\theta^* - \beta\,\theta^* + C\big)\sqrt{2\pi\sigma^{2*}}\big(\Phi(b) - \Phi(a)\big),$$

where $\Phi(.)$ represents the cumulative density function of the Normal(θ^*, σ^{2*}) distribution. (8)

The first task in Bayesian inference is compute the marginal distribution of each element in hyperparameter vector as

$$p(\boldsymbol{\psi}_k|\boldsymbol{y}) = \int p(\boldsymbol{\psi}|\boldsymbol{y})\,d\boldsymbol{\psi}_{-k}. \tag{9}$$

As stated in Blangiardo and Cameletti (2015), INLA exploits the assumptions of the model to produce a numerical approximation to the posteriors based on the Laplace approximation, so that the joint posterior of the hyperparameters can be approximated as

$$\begin{aligned} p(\boldsymbol{\psi}|\boldsymbol{y}) &= \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &\propto \frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{p(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi})p(\boldsymbol{\theta}|\boldsymbol{\psi})p(\boldsymbol{\psi})}{\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\psi})} \\ &=: \tilde{p}(\boldsymbol{\psi}|\boldsymbol{y}), \end{aligned} \tag{10}$$

where $\tilde{p}(\boldsymbol{\psi}|\boldsymbol{y})$ is the Gaussian approximation given by the Laplace method, and $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is the mode for given $\boldsymbol{\psi}$. The second task is compute posterior marginal distribution of each element of parameter vector as

$$\begin{aligned} p(\theta_i|\boldsymbol{y}) &= \int p(\theta_i,\boldsymbol{\psi}|\boldsymbol{y})\,d\boldsymbol{\psi} \\ &= \int p(\theta_i|\boldsymbol{\psi},\boldsymbol{y})\,p(\boldsymbol{\psi}|\boldsymbol{y})d\boldsymbol{\psi}. \end{aligned} \tag{11}$$

Rewrite the vector of parameters as $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$ and use again Laplace approximation, the joint posterior of parameters can be approximated as

$$\begin{aligned} p(\theta_i|\boldsymbol{\psi},\boldsymbol{y}) &= \frac{p((\theta_i,\boldsymbol{\theta}_{-i})|\boldsymbol{\psi},\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} \\ &\propto \frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})} \\ &\approx \left.\frac{p(\boldsymbol{\theta},\boldsymbol{\psi}|\boldsymbol{y})}{p(\boldsymbol{\theta}_{-i}|\theta_i,\boldsymbol{\psi},\boldsymbol{y})}\right|_{\boldsymbol{\theta}_{-i}=\boldsymbol{\theta}_{-i}{}^*(\theta_i,\boldsymbol{\psi})} \\ &=: \tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y}), \end{aligned} \tag{12}$$

where $\tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ is the Laplace Gaussian approximation and $\boldsymbol{\theta}_{-i}{}^*(\theta_i,\boldsymbol{\psi})$ is its mode. Once we get $\tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y})$ and $\tilde{p}(\boldsymbol{\psi}|\boldsymbol{y})$, the marginal posterior distributions as equation (11) then approximated by

$$\tilde{p}(\theta_i|\boldsymbol{y}) = \int \tilde{p}(\theta_i|\boldsymbol{\psi},\boldsymbol{y})\tilde{p}(\boldsymbol{\psi}|\boldsymbol{y})\;d\boldsymbol{\psi}, \tag{13}$$

and the integral can be solved numerically through a finite weighted sum

$$\tilde{p}(\theta_i|\boldsymbol{y}) \approx \sum_l \tilde{p}(\theta_i|\boldsymbol{\psi}^{(l)},\boldsymbol{y})\,\tilde{p}(\boldsymbol{\psi}^{(l)}|\boldsymbol{y})\Delta_l, \tag{14}$$

with a corresponding set of weights $\{\Delta_l\}$ from some integration points $\{\boldsymbol{\psi}^{(l)}\}$.

III. ANALYSIS AND RESULTS

3.1 Data set

We illustrate the INLA estimation results using the number of poor people data set in East Java province, Indonesia, year 2016 to be modeled using Bayes spatial and year 2012-2016 modeled using Bayes spatio-temporal. East Java is the second largest province in Indonesia (according to the number of population), with Surabaya as its capital city. Although Surabaya is the second biggest metropolis after Indonesia capital city, Jakarta, however, in province level, it has the biggest number of poor people in year 2014-2016 according to Statistics Indonesia data. Furthermore, Indonesia's economic crisis experienced in year 1998 produces a huge effect on the increase of the number of poor people to 49.50 (24.23%) million from, two years before in 1996, about 34.01 (17.47%) million people. This phenomenon interested us to discuss poverty and try to link various socioeconomics variables such as population density, expectation years of schooling and construction overpriced index.

East Java province has 29 districts and 9 municipalities, then we have 38 different regions with the average number of neighbor is 3.631579. Summary of the neighboring structure can be obtained using some of the libraries in R, such as maptools, spdep, and rgdal, while ploy2nb and nb2INLA respectively read and create the adjacency graph that will be used in R-INLA library package analysis. We concern in the number of poor people (in million) as the response variable, with range of the data is 272 and its minimum and maximum value are 8 and 280 respectively, thus our response fit to the generalized extreme value distribution with negative shape parameter k = -0.25222 which has finite right tail characteristics.

3.2 Bayes spatial model

Suppose y_i represents the number of poor people specified for ith region, $i = 1,2, \dots 38$, that we assume with generalized extreme value distribution. Then with identity link function we have

$$\eta_i = y_i = b_0 + \vartheta_i + \upsilon_i, \qquad (15)$$

with b_0 is the intercept and use Besag–York–Molliè (BYM) specification, thus ϑ_i is the spatially structured risk effect which assumed as in equation (2), and υ_i is the spatially unstructured risk effect assumed with exchangeable prior $\upsilon_i \sim \text{Normal}(0, \sigma_\upsilon^2)$.

We now prepare the BYM model and run INLA:

```
library(INLA)
formula <- y ~ 1 + f(ID, model="bym",
graph=JATIM.adj)
mod <-
inla(formula,family="gev",data=data,control.comput
e=list(dic=TRUE),control.family = list(gev.scale.xi =
0.03))
```

R-INLA library package can be downloaded from http://www.r-inla.org/ which provides documentation for the package and discussion forum. Once we had downloaded the package, library(INLA) starts the command for all analysis in INLA method. We define the regression formula, with 1 as the intercept, while f() states the model specification. We set the BYM regression model, model="bym", for each ID region with the JATIM.adj as the graph of its adjacency matrix. By default, R-INLA specified minimally informative priors on the log of the structured effect precision: $\log \tau_\vartheta \sim \log \text{Gamma}(1,\ 0.0005)$ and the unstructured effect precision: $\log \tau_\upsilon \sim \log \text{Gamma}(1,\ 0.0005)$. If the model has been specified, we can use the inla function to run INLA algorithm. In inla function, the formula is specified before using BYM specification, using the number of poor people as our data set with gev as the names of generalized extreme value for family distribution in inla function. For model comparisons we evaluate the Deviance Information Criteria (DIC) and use the scaling of shape parameter ξ_s is given by the argument gev.scale.xi with the default scaling is set to 0.01. The scaling of the shape parameter is purposed to improve the numerical calculation if ξ is small, so the shape parameter $\theta_2 (= \xi)$ will appear as θ_2 / ξ_s. Scaling may vary for different data characteristics while in our case we set the scale to 0.03.

In spatial analysis, it will be interesting to investigate the proportion of variance explained by the spatially structured component. The variance of the conditional autoregressive specification σ_ϑ^2 is not directly comparable with the variance of the marginal unstructured component σ_υ^2. Blangiardo and Cameletti (2015) obtain an estimate of the posterior marginal variance for the structured effect empirically through

$$s_\vartheta^2 = \frac{\sum_{i=1}^{n} (\vartheta_i - \bar{\vartheta})^2}{n - 1},$$

where $\bar{\vartheta}$ is the average of ϑ and then compare it to σ_υ^2, so we have the fraction of spatially structured effect as

$$\text{Frac}_{\text{ spatial}} = \frac{s_\vartheta^2}{s_\vartheta^2 + \sigma_\upsilon^2}.$$

In our data set, the proportion of spatial variation is about 29.88% where these numbers have small spatial effect to represents data variability. One of the excellences of Bayesian analysis is the flexibility on its prior distribution that may have a considerable impact on the results including on its spatial fraction. We improve the spatial fraction with define other assumption of its structured and unstructured effect prior distribution with the following argument

```
formula <- y ~ 1 + f(ID, model="bym",
graph=JATIM.adj,hyper=list(prec.unstruct=list(prior
="loggamma",param=c(1,0.0001)),prec.spatial=list(pri
or="loggamma",param=c(1,0.001))))
```

The above modification of the prior distribution is specified using the option hyper in the formula specification, note that now we use the minimally informative priors on the log of the structured effect precision as $\log\tau_{\vartheta} \sim \log\text{Gamma}(1,\ 0.001)$ and the unstructured effect precision as $\log\tau_{\upsilon} \sim \log\text{Gamma}(1,\ 0.0001)$. The improvement of its spatial fraction is simply by setting bigger precision (or equivalently smaller variance) for the spatially structured rather than the unstructured effect. By setting this new prior distribution for the structured and the unstructured effect, the spatial fraction increases to about 81.72%, therefore we will use this assumption for the BYM specification in Bayes spatial and Bayes spatio-temporal models.

With respect to Section 2.1, the parameters estimated by R-INLA are $\boldsymbol{\theta} = \{b_0, \boldsymbol{\vartheta}, \boldsymbol{\upsilon}\}$ and the hyperparameters are given by the precisions $\boldsymbol{\psi} = \{\tau_{\vartheta}, \tau_{\upsilon}\}$. The information about posterior mean, standard deviation and a 95% credibility interval for the fixed effect (its intercept, b_0) and the random effect (the structured and unstructured, $\boldsymbol{\vartheta}, \boldsymbol{\upsilon}$) can be obtained respectively with mod$summary.fixed and mod$summary.random. Our next interesting objective is to map the relative risk of number of poor people of each region compared to the whole of East Java province, so we need to calculate the marginal distribution of each province using mod$marginals.random, transform to its natural scale then calculate the posterior distribution as follows

```
m <- mod$marginals.random$ID[1:Nareas]
zeta <-
lapply(m,function(x)inla.emarginal(function(x) x,x))
```

m represents the marginal distribution of the linear combination of random effects as $\omega_i = \vartheta_i + \upsilon_i$ for each ID region which Nareas identifies the 38 regions in East Java, and inla.emarginal is the transformation function to its natural scale, while in our generalized extreme value the transformation function is identity, identified with inla.emarginal(function(x) x,x). Using built-in function in R-INLA, we calculate the posterior mean distribution with lapply function. The posterior mean of the intercept, b_0, implies a 134.235 (in million) increasing number of poor people, with 95% credibility interval ranging from 112.6586 to 156.1339. Figure 1(a) shows the map of the posterior mean of relative risk for number of poor people of each region ω_i, compared to the whole of East Java. The posterior means of the uncertainty can also be mapped and can provide useful information and visualized with $p(\omega_i > 0|\boldsymbol{y})$ using the built-in function inla.pmarginal, and the resulting map presented in Figure 1(c).

We investigate the relationship of the number of poor people using the additional socioeconomics covariates, so we define their impact in spatial regression formula (16)

$$\eta_i = y_i = b_0 + \vartheta_i + \upsilon_i + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}$$

where x_1, x_2 and x_3 for each region respectively describe population density, expectation years of schooling and construction overpriced index. Then we specify the formula.cov in inla function as below

```
formula.cov <- y ~ 1 + f(ID, model="bym",
graph=JATIM.adj,hyper=list(prec.unstruct=list(prior
="loggamma",param=c(1,0.0001)),prec.spatial=list(pri
or="loggamma",param=c(1,0.001)))) + x1 + x2 + x3
mod.cov <-
inla(formula.cov,family="gev",data=data,control.com
pute=list(dic=TRUE),control.family = list(gev.scale.xi
= 0.03))
```

Summary of INLA estimation for the fixed effects $\{b_0, x_1, x_2, x_3\}$ are presented in Table 1. On its natural scale, the increase of 1 unit in population density and construction overpriced index are associated respectively with an increase of around (in million) 13.74 and 10.70 in the number of poor people, while decrease around 46.40 for 1 unit increases in expectation years of schooling. Map of the residuals relative risks for each region and the uncertainty of their probability of exceeding 0 are presented in Figure 1(b, d).

Table 1. Summary Of Fixed Effects Estimation, Standard Deviation And 95% Credibility Interval

	Mean	Standard Deviation	2.5%	50%	97.5%
b_0	123.2651	10.7423	101.3054	123.3054	144.2790
x_1	13.7441	10.1164	-7.8815	14.0649	33.7265
x_2	-46.4029	10.7544	-67.4138	-46.4906	-25.0979
x_3	10.7012	10.5411	-9.9642	10.6671	31.4241

3.3 Bayes spatio-temporal models

This section we model the number of poor people of East Java, year 2012-2016 to build space-time relative risk of each region and then compare the resulted models from section 2.2.1 – 2.2.3 using DIC criteria. The classical non-parametric formulation as in equation (3) with identity link function can be written as

$$\eta_{it} = y_{it} = b_0 + \vartheta_i + \upsilon_i + (\beta + \delta_i)t. \quad (17)$$

In R-INLA model in equation (17) is specified through the following formula.1

```
formula.1<- y ~ 1 +
f(ID.area,model="bym",graph=Jatim.adj,hy
per=list(prec.unstruct=list(prior="logga
mma",param=c(1,0.0001)),prec.spatial=lis
t(prior="loggamma",param=c(1,0.001)))) +
f(ID.area1,year,model="iid") + year
model.inla.1 <-
inla(formula.1,family="gev",data=data,
control.predictor=list(compute=TRUE),
control.compute=list(dic=TRUE,cpo=TRUE),
control.family = list(gev.scale.xi =
0.03))
```

Model (17) includes the same spatial structured and unstructured components as in (16), with `year` is the global time effect (β). The linear combination of their random effects are $\omega_i = \vartheta_i + \upsilon_i$ using the BYM specification identified in f(ID.area,model="bym",…). Note that each function f() only assigned with one unique covariate in R-INLA, so that we have to define ID.area1 to duplicate ID.area used in previous specification. Therefore f(ID.area1,year,model="iid") are the random effects represents for space (ID.area1) and time (year)interaction (δ_i) and modeled using exchangeable prior Normal$(0,1/\tau_\delta)$ with $\log \tau_\delta \sim \log \text{Gamma}(1, 0.0005)$.

Using dynamic parametric trend, global time effect is replaced with the structured and unstructured temporal component, so that with identity link function equation (4) can be written as

$$\eta_{it} = y_{it} = b_0 + \vartheta_i + \upsilon_i + \rho_t + \varphi_t.$$

In `R-INLA` model in equation (18) is specified through the following `formula.2`

```
formula.2<- y ~ 1 +
f(ID.area,model="bym",graph=Jatim.adj,hy
per=list(prec.unstruct=list(prior="logga
mma",param=c(1,0.0001)),prec.spatial=lis
t(prior="loggamma",param=c(1,0.001)))) +
f(ID.year,model="rw1") +
f(ID.year1,model="iid")
model.inla.2 <-
inla(formula.2,family="gev",data=data,
control.predictor=list(compute=TRUE),
control.compute=list(dic=TRUE,cpo=TRUE),
control.family = list(gev.scale.xi =
0.03))
```

The structured temporal trends (ρ_t) modeled using random walk of order one as in equation (5) through `f(ID.year,model="rw1")`, and `f(ID.year1,model="iid")` is the unstructured temporal trends with exchangeable prior Normal$(0,1/\tau_\varphi)$ with $\log \tau_\varphi \sim \log \text{Gamma}(1, 0.0005)$.

In space-time interaction model, we define specifically the interaction between the spatially unstructured random effects (υ_i) and the unstructured temporal trends (φ_t), so that according to equation (6) with identity link function, the linear predictor can be written as

$$\eta_{it} = y_{it} = b_0 + \vartheta_i + \upsilon_i + \rho_t + \varphi_t + \gamma_{it}. \quad (19)$$

In R-INLA model in equation (18) is specified through the following formula.3

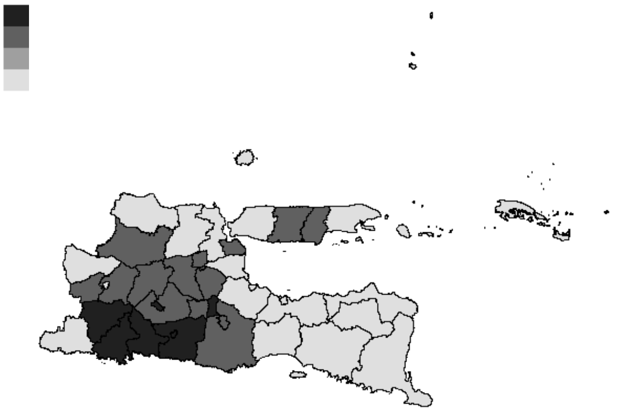
```
formula.3<- y ~ 1 +
f(ID.area,model="bym",graph=Jatim.adj,hy
per=list(prec.unstruct=list(prior="logga
mma",param=c(1,0.0001)),prec.spatial=lis
t(prior="loggamma",param=c(1,0.001)))) +
f(ID.year,model="rw1") +
f(ID.year1,model="iid") +
f(ID.area.year,model="iid")
lcs = inla.make.lincombs(ID.year =
inla(formula.3,family="gev",data=data,
control.predictor=list(compute=TRUE),
control.compute=list(dic=TRUE,cpo=TRUE),
lincomb=lcs,control.inla =
list(lincomb.derived.only=TRUE),control.
family = list(gev.scale.xi = 0.03))
```

(18)

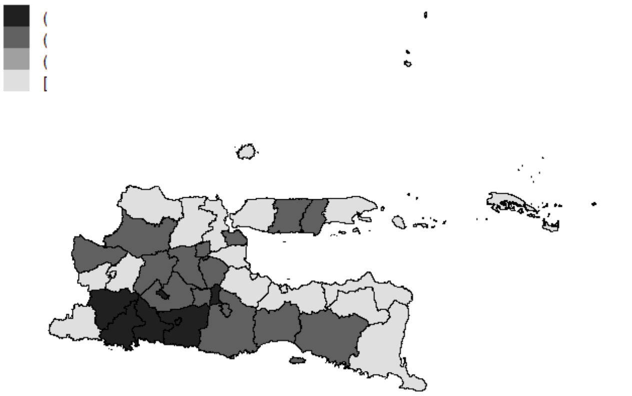
Model (19) has the same structured and unstructured temporal trends as in (18), while the space-time interaction specified using f(ID.area.year,model="iid") with follows a Gaussian distribution, i.e. $\gamma_{it} \sim \text{Normal}(0, 1/\tau_\gamma)$ with $\log \tau_\gamma \sim \log \text{Gamma}(1,\ 0.0005)$ and the structure matrix defined as in Section 2.2.3. The two temporal effects $\{\rho_t, \varphi_t\}$ can be combined through a linear combination using inla.make.lincombs before running the model, in this case we store in lcs variable. The linear combination is obtained using the 5 elements of two diagonal matrices, one for the structured and one for the unstructured temporal trend parameters.

Table 2. Results Of Spatio-Temporal Models

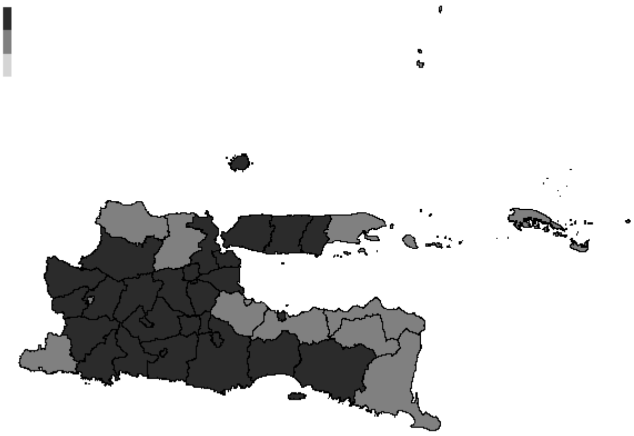
Model	DIC	p_D
1	2120.101	2.159
2	2120.921	2.777
3	2121.062	2.858



(a) Marginal posterior mean distribution of relative risks of number of poor people ω_i in model (15)



(b) Marginal posterior mean distribution of residual relative risks of number of poor people ω_i with socio-economic covariates in model (16)



(c) The uncertainty of relative risks probability $p(\omega_i > 0|\boldsymbol{y})$ in model (15)

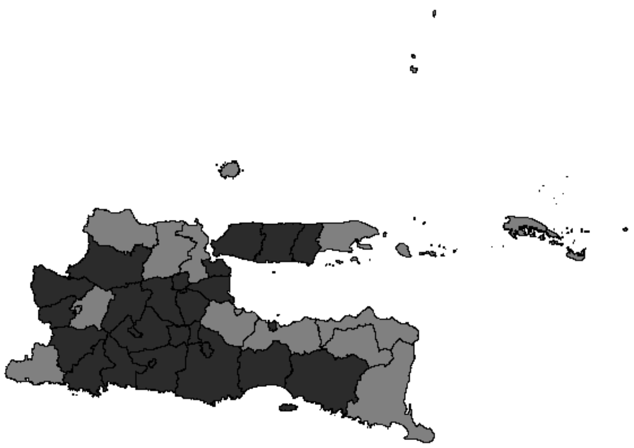
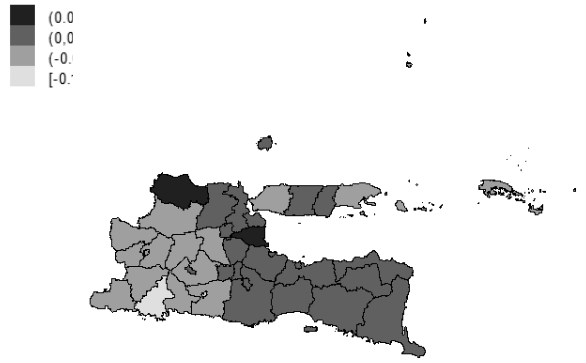
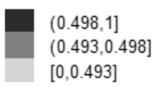


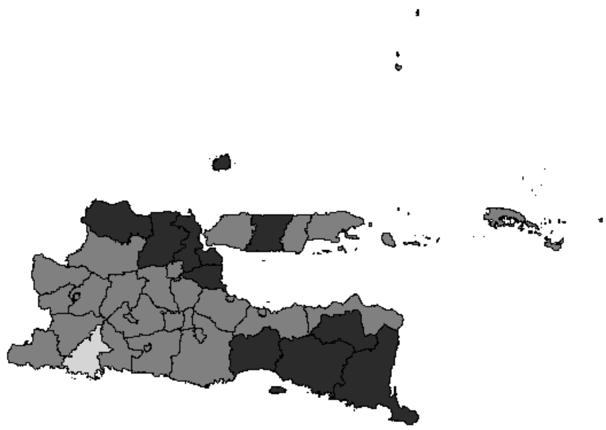
Figure 1. Spatial mapping results and posterior probabilities for Bayes spatial model

Table 2 presents the results of spatio-temporal models defined in equation (17) – (19) with DIC criteria (smaller is better) and the numbers of effective parameters (p_D) that represents the model complexity. Classical parametric trend improves the model fit with smaller DIC, suggesting that the added complexity from the number of parameters to be estimated (like in non-parametric trend models) does not guarantee will result in the most fit model, therefor we will focus on classical parametric trend results. In this model the intercept posterior mean is 135.6424 (in million) ranging from (115.2128, 156.4254) on it 95% credibility interval with standard deviation is 10.4910. The negative posterior mean of global temporal trend is $\beta = -1.6766$, shows decreasing linear trend in number of poor people in 5 years observations. Spatial relative risk also mapped in Figure 2(a) while the uncertainty of $p(\omega_i > 0|\boldsymbol{y})$ is presented in Figure 2(b).



(a) Marginal posterior mean distribution of relative risks of number of poor people ω_i in model (17)





(b) The uncertainty of relative risks probability $p(\omega_i > 0|\boldsymbol{y})$ in model (17)

Figure 2. Spatial mapping results and posterior probabilities for Bayes classical parametric trend

IV. DISCUSSIONS

This paper briefly describes INLA algorithm to estimate marginal posterior mean for parameters and hyperparameters for Bayes spatial to spatio-temporal models. We use the number of poor people data set that fit to generalized extreme value distribution to illustrate the INLA estimation results. The INLA produced great computational benefits rather than MCMC method in solving big-n problem that covers random and fixed effects to every specific region and time on its spatio-temporal analysis. We model the number of poor people in East Java province, Indonesia, year 2016 using Bayes spatial with Besag–York–Mollié (BYM) specification. The spatial pattern of this model is described with spatial fraction derived from spatially structured to spatially unstructured effect, which produce 81.72% of data variability that can be described with its spatial relationship. We improve the spatial fraction using different prior assumption of hyperparameters used in R-INLA. This flexibility of choosing the prior distribution assumption is one of the excellences of Bayesian modeling and at the same time being an issue for the accuracy of prior assumption used against the resulting estimates. Therefore, it is necessary to set carefully the prior is being used and perform sensitivity analysis to investigate how the prior influences the estimators.

We add the Bayes spatial model using socioeconomics covariates and explore how its effects to the response variable. From the result, we can conclude that expectation years of schooling have greatest effect in determining the number of poor people in East Java, the bigger the expectation years of schooling the less the number of poor people. In other hand, population density and construction overpriced index have negative influence, so that the bigger the values the more the number of poor people. Before analysis and as the second most densely populated province, we thought that population density will have the greatest effect in East Java poor people numbers, however, it is surprising that expectation years of schooling have more than three times effects rather than population density and almost five times greater than construction overpriced index. Therefor in this case, rather than density and infrastructure, the role of education is the most influential on the development of welfare population.

Using the same BYM specification and assumptions, we again model the number of poor people year 2012-2016 using classical, dynamic and space-time interaction of spatio-temporal models. The best fit model accordance to smallest DIC criteria is classical model with parametric trend. Rather than the nonparametric trend models, classical Bayes Spatio-temporal is the simplest model which only assume the global linear trend (β), so that with (very short) only five years observation it is very reasonable that the simplest classical model is the best fit for our data set. If we have longer observation, it is possible that the nonparametric models can have the varying temporal trend and may also be able to capture the space-time interaction. Note that in space-time interaction model, we only use the type I interaction, which interact the unstructured temporal effects (φ_t) and spatially unstructured effects (υ_i). Modelling the combination of spatial and temporal interaction also interesting, such as type II (temporal structured (ρ_t) and spatially unstructured (υ_i) interaction), type III (unstructured temporal (φ_t) and spatially structured (ϑ_i) interaction) and most complex interaction type IV (temporal structured (ρ_t) and spatially structured (ϑ_i)) (Blangiardo and Cameletti (2015)). However, when the covariates are available, it could be interesting to study their effect on ecological regression (regression using covariates) to its spatio-temporal analysis.

V. AKCNOWLEDGEMENTS

The authors want to thank LPDP (Lembaga Pengelola Dana Pendidikan) and Kemenristek DIKTI for supporting BUDI (Beasiswa Unggulan Dosen Indonesia) Scholarship for the first author, and Bina Nusantara University for supporting the first author with various publication development program and full affiliation.

VI. REFERENCES

[1] S. Banarjee, B. P. Carlin and A. E. Gelfand, “Hierarchical modelling and analysis for spatial data”, 2nd ed. US: CRC Press; 2015.

[2] M. Blangiardo and M. Cameletti, “Spatial and spatio-temporal Bayesian models with R-INLA”, 1st ed. UK: John Wiley & Sons; 2015.

[3] J. R. P. D. Carvalho, M. N. Alan and E. B. A. M. José. 2016. Spatio-temporal modeling of data imputation for daily rainfall series in homogeneous zones. Revista Brasileira de Meteorologia. 31(2):196-201. DOI: 10.1590/0102-778631220150025.

[4] J. C. Aravena and B. H. Luckman. 2009. Spatio-temporal rainfall patterns in southern south america. International Journal of Climatology. 29(14):2106-2120. DOI: 10.1002/joc.1761.

[5] J. D. Yanosky, C. J. Paciorek, F. Laden, J. E. Hart, R. C. Puett, D. Liao and H. H. Suh. 2014. Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological

predictors. Environmental Health. 13(63). DOI: 10.1186/1476-069X-13-63.

[6] R. Ye, K. Shan, H. Gao, R. Zhang, W. Xiaong, Y. Wang and X. Qian. 2014. Spatio-temporal distribution patterns in environmental factors, chlorophyll-a and microcystins in a large shallow lake. International Journal of Environmental Research and Public Health. 11(5):5155-69. DOI: 10.3390/ijerph110505155.

[7] J. J. Abellan, S. Richardson and N. Best. 2008. Use of space-time models to investigate thestability of patterns of disease. Environmental Health Perspective. 116(8):1111–9. DOI: 10.1289/ehp.10814.

[8] Z. Ma, X. Hu, A. M. Sayer, R. Levy, Q. Zhang, Y. Xue, S. Tong, J. Bi, L. Huang L and Y. Liu. 2016. Satellite-based spatiotemporal trends in PM2.5 concentrations: china, 2004–2013. Environmental Health Perspectives. 124(2):184-92. DOI: 10.1289/ehp.1409481.

[9] U. Dieckman, T. Herben and R. Law. 1999. Spatio-temporal processes in ecological communities. CWI Quarterly. 12(3&4):213 – 38.

[10] F. I. Korennoy, V. M. Gulekin, J. B. Malone, Cn. Mores, S. A. Dudnikov and M. A. Stevenson. 2014. Spatio-temporal modeling of the African swine fever epidemic in the Russian Federation, 2007–2012. Spatial and Spatio-temporal Epidemiology. 11: 135-41. DOI: 10.1016/j.sste.2014.04.002.

[11] M. Nazia, M. Ali, Md. Jakariya, Q. Nahar, M. Yunus and M. Emch. 2018. Spatial and population drivers of persistent cholera transmission in rural Bangladesh: Implications for vaccine and intervention targeting. Spatial and Spatio-temporal Epidemiology. 24: 1–9. DOI: 10.1016/j.sste.2017.09.001.

[12] D. Kaplan, "Bayesian Statistics for the Social Sciences", 1st ed. New York: The Guilford Press; 2014.

[13] S. Jackman, "Bayesian Analysis for the Social Sciences", 1st ed. United Kingdom: Wiley; 2009.

[14] G. Baio, A. Berardi and A. Heath, "Bayesian Cost-Effectiveness Analysis with the R package BCEA", 1st ed. Switzerland: Springer International Publishing; 2017.

[15] M. Pirani, J. Gulliver, G. W. Fuller and M. Blangiardo. 2014. Bayesian spatiotemporal modelling for the assessment of short-term exposure to particle pollution in urban areas. Journal of Exposure Science and Environmental Epidemiology. 24: 319–27. oi:10.1038/jes.2013.85.

[16] G. Casella and E. George. 1992. Explaining the Gibbs sampler. The American Statistician. 46(3): 167–74.

[17] W. Gilks, S. Richardson and D Spiegelhalter, "Markov Chain Monte Carlo in Practice", 1st ed. Florida: Chapman & Hall/CRC; 1996.

[18] J. Besag, J. York and A. Mollie. 1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics. 43(1): 1–59.

[19] S. Banerjee., B. Carlin and A. Gelfand, "Hierarchical modeling and analysis for spatial data", 2nd Ed. US: CRC Press. 2015.

[20] P. Diggle and J. P. Ribeiro, "Model-based Geostatistics", 1st ed. New York: Springer-Verlag; 2007.

[21] N. Cressie and C. Wikle, "Statistics for spatio-temporal data", 1st ed. New Jersey: Wiley; 2011.

[22] I. Ntzoufras, "Bayesian modeling using WinBUGS", 1st ed. United States: John Wiley and Sons; 2009.

[23] H. Rue, S. Martino and N. Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of Royal Statistical Society B. 71(2):1–35.

[24] S. Martin and H. 2010. Rue. Implementing approximate bayesian inference using integrated nested laplace approximation: a manual for the inla program. Available from: http://www.math.ntnu.no/hrue/ GMRFsim/manual.pdf.

[25] R. Ruiz-Cárdenas, E. Krainski and H. Rue. 2012. Direct fitting of dynamic models using integrated nested laplace approximations INLA. Computational Statistics & Data Analysis. 56(6):1808–28.

[26] S. Martino, K. Aas, O. Lindqvist, L. Neef and H. Rue. 2011. Estimating stochastic volatility models using integrated nested laplace approximations. The Europian Journal of Finance 17(7):487–503. DOI: 10.1080/1351847X.2010.495475.

[27] D. Simpson, J. B. Illian, F. Lindgren, S. H. Sørbye SH and H. Rue. 2016. Going off grid: computationally efficient inference for log-Gaussian Cox processes. Biometrika. 1–22. DOI: 10.1093/biomet/asv064.

[28] A. Cosandey-Godi, E. T. Krainski, B. Worm anf J. M. Flemming. 2015. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Canadian Journal of Fisheris and Aquatic Science. 72(2): 1–12. DOI.org/10.1139/cjfas-2014-0159.

[29] M. Cameletti, F. Lindgren, D. Simpson and H. Rue. 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. Advances in Statistical Analysis. 97(2): 109–31.

[30] A. L. Papiola, A. Riebler, A. Amaral-Turkman, R. São-João, C. Ribeiro, C. Geraldes and A. Miranda. 2013. Stomach cancer incidence in Southern Portugal 1998-2006: a spatio-temporal analysis. Biometrical Journal. 56(3):403-15. DOI: 10.1002/bimj.201200264.

[31] B. Mahmoudian and M. Mohammadzadeh. 2014. A spatio-temporal dynamic regression model for extreme wind speeds. Extremes. 17(2):221–45.

[32] S. Ghosh and B. K. Mallick. 2011. A hierarchical Bayesian spatio-temporal model for extreme precipitation events. Environmetrics. Research Article: Wiley Online Library. 22(2):192-204. DOI: 10.1002/env.1043.

[33] S. Pindado, C. Pindado C and J. Cubas J. 2017. Fréchet distribution applied to salary incomes in Spain from 1999 to 2014. An engineering approach to changes in salaries' distribution. Economies. 2017; 5(2): 14. DOI:10.3390/economies5020014.

[34] R. A. Fisher and L. H. C. Tippett. 1928. Limiting forms of the frequency distributions of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society. 24(2):180-90. DOI: 10.1017/S0305004100015681.

[35] H. Rue H and L. Held, "Gaussian Markov random fields", Theory and applications. Chapman & Hall; 2005.

[36] J. Besag. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). Journal of Royal Statistical Society, Series B. 36(2):192–236.

[37] C. Herrmann, S. Ess1, B. Thürlimann, N. Probst-Hensch and P. Vounatsou. 2015. 40 years of progress in female cancer death risk: a Bayesian spatio-temporal mapping analysis in Switzerland. BMC Cancer. 15:666-675. DOI: 10.1186/s12885-015-1660-8.

[38] T. Gneiting T, M. G. Genton and P. Guttorp, "Statistical methods for spatio-temporal systems", Edited by Finkenstädt B, Held L, Isham V, in Statistical Methods for Spatio-temporal Systems. CRC Press, Chapmann and Hall; 2006:151–75.

[39] D. Clayton, "Generalised linear mixed models", Edited by Gilks W, Richardson S, Spiegelhalter D in Markov Chain Monte Carlo in Practice. Chapman & Hall; 1996:275–301.

[40] L. Knorr-Held. 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine. 19(17–18):2555–67.

[41] L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi and M. Songini. 1995. Bayesian analysis of space-time variation in disease risk. Statistics in Medicine. 14(21–22):2433–43.

[42] http://www.r-inla.org