Monitoring the Underlying Probability of Daily Rainfall Occurrence in West Java over Space and Time by Means: A Bayesian Spatiotemporal Model Approach

Jaya, I. G. N. M.,[1,2*] Ruchjana, B. B.,[3] Abdullah, A. S.[4] and Andriyana, Y.[5]
[1]Statistics Department, Universitas Padjadjaran, Bandung, Indonesia, E-mail: mindra@unpad.ac.id
[2]Faculty Spatial Science, Groningen University, Groningen, Netherlands
[3]Mathematics Department, Universitas Padjadjaran, Bandung, Indonesia
[4]Computer Science, Universitas Padjadjaran, Bandung, Indonesia
[5]Statistics Department, Universitas Padjadjaran, Bandung, Indonesia
[*]*Corresponding author*

Abstract

Monitoring the underlying probability of daily rainfall occurrence is needed for many applications, such as in agriculture, hydrology and ecology, epidemiology, maritime and tourism and society in general. We use Climate Hazard InfraRed Precipitation with Station Data (CHIRPS) to obtain the estimated probability of daily rainfall occurrence. The model has been used based on Kitagawa (1987). We used the Kitagawa model with introducing the spatial and spatiotemporal dependencies in the Binomial smoothing model using Bayesian spatiotemporal model approach by means of the Integrated Nested Laplace Approach (INLA) and applied this model for estimating the underlying probability of daily rainfall occurrence in West Java, Indonesia. Eight different model were tested and model 8 was found as the best model based on Bayesian and classical criterions. We found the circular temporal trend of random walk order one and spatiotemporal interaction type IV are the two most important components in explaining the change of probability of daily rainfall occurrence. The daily precipitation chances reached 90% from November up to February for every year.

1. Introduction

The accurate estimation of the probabilistic prediction for the daily rainfall is needed for many applications related to agriculture, hydrology and ecology, epidemiology, maritime, tourism and society in general (Roebber and Bosart, 1996, Wilson and Toumi, 2005, Singh et al., 2012, Thach and Canh, 2013 and Jayasinghe et al., 2015). In a tropical country such as Indonesia which is highly dependent on rain for its food production, tourism and hydropower generation, the variability of rainfall hence becomes a strong limiting factor (Perera et al., 2002). However, the rainfall density data is not easily to be obtained because of the limitation of the rainfall monitoring station. This problem can be solved by mining data from the satellites which provide rainfall information. One of the satellite data that can be utilized to obtain data of rainfall intensity is the Climate Hazard InfraRed Precipitation with Station Data (CHIRPS). CHIRPS is rainfall data that utilizes infrared waves and rain stations with 0.05^o spatial resolution (per pixel) to estimate the value of continuous rainfall in a region as well as for drought trend analysis of the area. Another advantage of using CHIRPS data is that it can give both daily and monthly rainfall. We can optimize the utilization of CHIRPS data by transforming data into frequency data and then smoothing it to see the pattern of rainfall frequency in one year.

This information can be used to predict on which day and month the frequency of rainfall can be high and estimate the highest and lowest probability value of daily rainfall occurrence (Singh et al., 2012 and Kumar and Bhardwaj, 2015). Markov chain models play an important role in modeling rainfall data especially wet-dry sequence (Smith, 1987 and Chowdhury et al., 2017). Two of the most interesting characteristics of the Markov chains models is that i) it is easy to apply in modeling seasonality and ii) the parameter models can be estimated and tested using effective statistical inference procedure (Coe and Stern, 1984). The Bernoulli process is a part of the Markov Chain model.

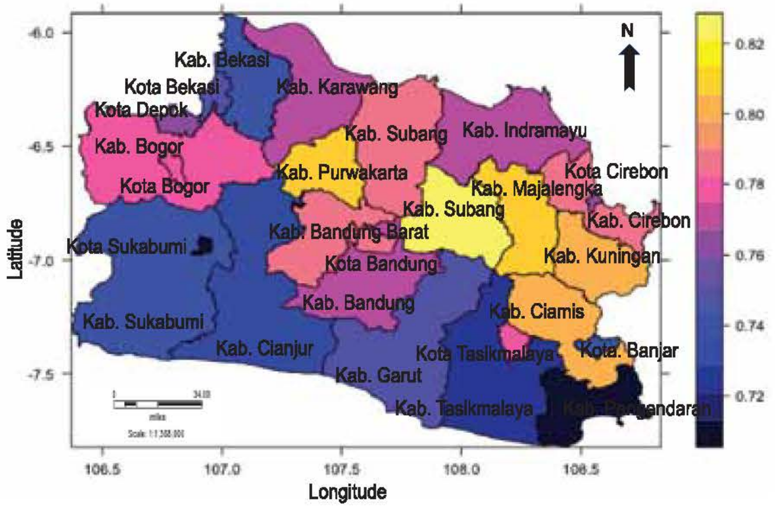


Figure 1: The empirical probability of daily rainfall occurrence 2010-2015 in West Java, Indonesia

Kitagawa (1987) introduced the method of smoothing time series Binomial data to estimate the daily probability of the occurrence rainfall in Tokyo (19803-1984) at calendar day $i = 1,2, \dots ,365$, which was assumed to be gradually changing with time. The model is developed based on circular random walk model. Here, we expand the Kitagawa model to estimate the underlying probability of daily rainfall occurrence data in West Java. In total, West Java has 27 small districts. West Java province is one of the worst affected areas by high rainfall and thus flooding is still makes it to one of the major problems that is being faced by this province of Indonesia. Figure 1 displays the choropleth map of empirical probability of daily rainfall occurrence 2000-2015 in West Java, Indonesia. The Figure 1 indicates that there is spatial dependence where there is any spatial cluster. Spatial cluster occurs when a collection of spatial units that have similar characteristics (Neethu and Surendran, 2013). The spatial dependence in rainfall occurrences can be caused by wind movement. The regions in southern of West Java such as Sumedang, Purwakarta, Ciamis, Kuningan, and Majalengka are the five districts which have the highest wet periods compared to other districts in West Java. Those can be categorized as a cluster of districts with high rainfall occurrence.

Wind direction can affect the movement of clouds containing rainwater so that adjacent locations tend to have the same daily rainfall occurrence (Johansson and Chen, 2003). Using Moran's Index statistics, we found the spatial autocorrelation 0.205 (p-value=0.042) which indicates there is a significant spatial autocorrelation in the data. Because of this phenomenon, considering the spatial and temporal dependency are important in modeling spatiotemporal data to minimize bias and increase precision estimates (Stauffer et al., 2017 and Jaya et al., 2018a, 2018b). In this study, we estimated the probability of daily rainfall occurrence for all districts simultaneously using Kitagawa model. We introduced the spatiotemporal model to accommodate the spatial and temporal dependency and heterogeneities in order to smooth parameter estimates due to the random error sampling variation. The novelty of this paper is to develop models by accommodating not only the temporal but also spatial and temporal components simultaneously.

To consider the spatial dependence in estimating the underlying probability of rainfall on a given calendar day we introduced intrinsic Conditional Auto-Regressive ($iCAR$) model by Besag et al., (1974) and circular random walk (RW) model for the temporal autocorrelation for smoothing time series of Binomial data. We also considered the four types spatiotemporal interaction in estimating the underlying probability of daily rainfall occurrence (Knorr-Held, 2000 and Bivand et al., 2013).

Instead of MCMC approximation, we considered using Integrated Nested Laplace Approximation (INLA) to estimate the parameters model. MCMC methods are widely used for Bayesian inference. However their computational burden lead to serious problems for modelling big dataset and with complex specifications. INLA is a new Bayesian technique which provides accurate and fast result for any complex model (e.g., spatial and spatiotemporal models). INLA employs deterministic algorithm for Bayesian inference (Rue and Martino, 2009, Blangiardo et al., 2013 and Jaya et al., 2016, 2017). The structure of the paper is organized as follows: Section 2 comprises of the description of the spatiotemporal model to be fitted with INLA, while Section 3 gives a summary of the INLA method and the resulting analysis of rainfall data in West Java. Section 4 gives the discussion and final conclusion.

2. Fit Model Daily Rainfall Data by means of INLA

All the occurrences of rainfall which is over 1 mm in the West Java Province area for each calendar year during a 16 years (2010-2015) is mined from CHIRPS data. West Java has 27 districts each of which each has a different underlying probability of daily rainfall occurrence. It is important to estimate the underlying probability at district i at time point t, p_{it} of rainfall for calendar day which is assumed to change gradually over time. For each day $t = 1,\dots,365$ of the year we have the number of days that rained over district i denoted by y_{it} and the number of days that were observed for district i denoted by n_{it} . We considered a Bernoulli observational model for binary responses, denoted by $B(p_{it})$ a Bernoulli distribution with probability p_{it} for 1 and $1-p_{it}$ for 0. The most prominent regression model in this framework is Logit model (Martins, 2014). The model is given by a conditionally independent of Binomial likelihood function:

$$y_{it}|\eta_{it} \sim Bin(n_{it}, p_{it}), t = 1,2,\dots,365$$

Equation 1

with logit link function:

$$p_{it} = \frac{\exp(\eta_{it})}{1+\exp(\eta_{it})}$$

Equation 1

The daily rainfall occurrence has space and time dimensions so that the regression model should accommodate the spatially and temporally dependencies and heterogeneities or structured and unstructured components. The modelling framework is presented in Figure 2.

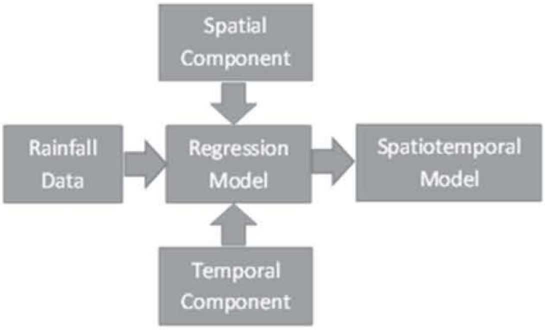


Figure 2: The modelling framework

Here, we assume the daily rainfall occurrence follows this specification:

$$\eta_{it} = \beta Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t + \delta_{it}$$

Equation 2

where β denotes the fixed effect temporal linear trend, where $\{v_i, \gamma_i, \omega_t, \phi_t, \varphi_t, \delta_{it}\}$ denotes the random effects components. v_i and γ_i are spatially structured and unstructured components respectively, while ω_t and ϕ_t denote temporally structured and unstructured components and δ_{it} is spatiotemporal interaction component. To estimate the model in Equation (3) we considered the Bayesian approach. The first component of Bayesian method is the likelihood function. Under the independent assumption, the likelihood function of Binomial process is defined by:

$$p(\mathbf{y}|\boldsymbol{\theta},\psi) = \prod_{i=1}^{n}\prod_{t=1}^{T}\binom{n_{it}}{y_{it}} p_{it}^{y_{it}}(1-p_{it})^{n_{it}-y_{it}}$$

Equation 3

where $\boldsymbol{\theta} = (\beta, v, \gamma, \omega, \phi, \varphi, \delta)'$ is the vector parameter model and $\boldsymbol{\psi} = (\psi_v, \psi_\gamma, \psi_\omega, \psi_\phi, \psi_\varphi, \psi_\delta)'$ is its hyperparameters. The second component is the prior distribution of each parameter. Spatially structured component is modelled by means $iCAR$ prior with precision ψ_v. The model usually used to present the structured spatial effect is the Besag, York and Mollié (BYM) model (Besag, 1974 and Besag et al., 1991). The prior can be written as:

$$v_i|\boldsymbol{v}_{-i}, \psi_v, \boldsymbol{W} \sim N\left(\frac{\sum_{j=1}^{n} w_{ij}v_j}{\sum_{i=1}^{n} w_{ij}}, \frac{1}{\psi_v \sum_{i=1}^{n} w_{ij}}\right), \forall\, t\,; i = 1,2,\dots,n$$

Equation 4

where $\boldsymbol{v}_{-i}$ denotes all the elements in $\boldsymbol{\omega}$ except the i^{th} element, $\boldsymbol{W} = (w_{ij})$ is the "neighborhood" matrix with $w_{ij} = 1$ if i and j being adjacent and $w_{ij} = 0$ otherwise. The spatially unstructured component is modeled using exchangeable priors $N(0, \psi_\gamma^{-1})$. The temporally structured, ϕ_t follows a circular random walk 1 (RW1) model with precision ψ_ϕ. The RW1 model is defined by:

$$\phi_{t+1} - \phi_t|\psi_\phi \sim N\left(0, \frac{1}{\tau_\phi}\right), \forall\, i;\ t = 1,2,..,T-1$$

Equation 5

The fact that we used a circular model here means that in this case, ϕ_1 is a neighbour of ϕ_{365} , since it makes sense to assume that the last day of the year has a similar effect when compared with the first day. We can expand the model by including seasonal effect ζ_t with the prior:

$$\zeta_t + \zeta_{t+1} + \cdots + \zeta_{t+m-1} | \psi_\zeta \sim N\left(0, \frac{1}{\psi_\zeta}\right), \forall\, i, t \\ = 1,2,..,T - m + 1,$$

Equation 6

where m denotes seasonal length. Similar to spatial unstructured component, the temporally unstructured (φ) is also modelled using exchangeable prior $N(0, \psi_\varphi^{-1})$. The spatial interaction component (δ) can be modelled using four type interaction (Knorr-Held, 2000) (Table 1).

Table 1: Four types spatiotemporal interaction

Type of Interaction	Interacting Parameters
I	γ_i and φ_t
II	γ_i and ϕ_t
III	φ_t and v_i
IV	v_i and ϕ_t

The third component is the posterior distribution which can be expressed as:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi}) \times p(\boldsymbol{\theta} | \boldsymbol{\tau}) \times p(\boldsymbol{\psi})$$

Equation 7

where "∝ " is called "proportional to". The posterior distribution in Equation (8) can be computed by means of Integrated Nested Laplace Approach (INLA). The regression model in INLA is defined as a function of a structured additive predictors $E(y_{it}) = \eta_{it}$ through a link function g(.):

$$\eta_{it} = \alpha + \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{l=1}^{L} f_l(z_{lit})$$

Equation 8

where α denotes intercept; β_k is a coefficient regression for covariate X_k and $f_l(.)$ is a function of random component z_l (Blangiardo and Cameletti, 2015). There are three stages to implement INLA for the Bayesian inference. The first stage defines the distribution of observational model $\boldsymbol{\pi(y|\theta)}$; the second stage is the latent Gaussian field (GMRF), $\boldsymbol{\pi}(\boldsymbol{\theta}|\boldsymbol{\psi}) \sim N(\boldsymbol{\mu_\psi}, \boldsymbol{Q_\psi^{-1}})$, and the third stage is the hyperparameter $\boldsymbol{\pi}(\boldsymbol{\psi})$ controlling parameters model. Here, **y** denotes the number of occurrences of daily rainfall since (2000-15) and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{v}, \boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\delta})'$ is the random vector parameter of the model (3) which have Gaussian priors; and $\boldsymbol{Q_\psi} = \boldsymbol{\Sigma}^{-1}$ denotes the precision matrix. Combining the three levels the joint posterior gives (Blangiardo et al., 2013).

$$\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\psi} | \boldsymbol{y}) = \boldsymbol{\pi}(\boldsymbol{\psi}) N(\boldsymbol{\mu_\psi}, \boldsymbol{Q_\psi^{-1}}) \prod_{i=1}^{N} \boldsymbol{\pi}(\boldsymbol{y_i} | \boldsymbol{\theta_i})$$

Equation 9

The precision matrix $\boldsymbol{Q_\psi}$ may be specified as a function of the structured matrix $\boldsymbol{R}$,

$$\boldsymbol{Q_\psi} = \tau \boldsymbol{R},$$

Equation 10

where τ is a hyperparameter and:

$$\boldsymbol{R} = \begin{cases} N_i & if\ i = j \\ -1 & if\ i \sim j \\ 0 \text{ otherwise} \end{cases}$$

Equation 11

The main objective is to estimate the marginal posterior distribution of all components of the GMRF:

$$\pi(\boldsymbol{\theta}_i | \boldsymbol{y}) = \int_{\boldsymbol{\psi}} \pi(\boldsymbol{\theta}_i | \boldsymbol{\psi}, \boldsymbol{y})\, \pi(\boldsymbol{\psi} | \boldsymbol{y}) d\boldsymbol{\psi},$$

Equation 12

where $\boldsymbol{\psi} = (\psi_v, \psi_\gamma, \psi_\omega, \psi_{\boldsymbol{\phi}}, \psi_\varphi, \psi_\delta)$. Thus, we need to find $\pi(\boldsymbol{\theta}_i | \boldsymbol{\psi}, \boldsymbol{y})$ and $\pi(\boldsymbol{\psi} | \boldsymbol{y})$. The first task (i) is find the marginal distribution of the hyperparameter (Blangiardo et al, 2013). The marginal posterior density $\pi(\boldsymbol{\psi} | \boldsymbol{y})$ of the hyperparameters $\boldsymbol{\psi}$ can be approximated using Laplace Approximation:

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

Equation 13

where $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{\psi}, \boldsymbol{y})$ is the Gaussian approximation (Rue et al., 2009) of $\pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \boldsymbol{y})$ and $\boldsymbol{\theta}^*(\boldsymbol{\psi})$ is the mode of $\pi(\boldsymbol{\theta}|\boldsymbol{\psi}, \boldsymbol{y})$. The second task (ii) is computing the $\pi(\boldsymbol{\theta}_{\mathrm{i}}|\boldsymbol{\psi}, \mathbf{y})$

$$\pi(\boldsymbol{\theta}_{\mathrm{i}}|\boldsymbol{\psi}, \mathbf{y}) = \frac{\pi\big((\boldsymbol{\theta}_{\mathrm{i}}, \boldsymbol{\theta}_{-\mathrm{i}})\big|\boldsymbol{\psi}, \mathbf{y}\big)}{\pi(\boldsymbol{\theta}_{-\mathrm{i}}|\boldsymbol{\theta}_{\mathrm{i}}, \boldsymbol{\psi}, \mathbf{y})} \approx \left.\frac{\pi(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y})}{\tilde{\pi}(\boldsymbol{\theta}_{-\mathrm{i}}|\Phi_{\mathrm{i}}, \boldsymbol{\psi}, \mathbf{y})}\right|_{\boldsymbol{\Phi}_{-\mathrm{i}}=\boldsymbol{\Phi}^*_{-i}(\Phi_{\mathrm{i}},\boldsymbol{\psi})}$$

Equation 14

where $\tilde{\pi}(\boldsymbol{\theta}_{-\mathrm{i}}|\theta_{\mathrm{i}}, \boldsymbol{\psi}, \mathbf{y})$ is the Gaussian approximation of $\pi(\theta_{-\mathrm{i}}|\Phi_{\mathrm{i}}, \boldsymbol{\psi}, \mathbf{y})$ and $\theta^*_{-i}(\theta_{\mathrm{i}}, \boldsymbol{\psi})$ is the mode of $\pi(\theta_{-\mathrm{i}}|\theta_{\mathrm{i}}, \boldsymbol{\psi}, \mathbf{y})$. The marginal posterior mean or median in Equation (15) is usually used to present the parameters estimate of the Model (3). The following prior of hyperparameters distributions are specified as under:

$$\beta \sim N(0, 10^5)$$
$$(\psi_v, \psi_\gamma, \psi_\omega, \psi_\phi, \psi_\varphi, \psi_\delta) \sim \text{Gamma}(1, 0.0005)$$

The hyper prior distributions given on those parameters are independent and weakly informative which is means the inference of parameters model are totally based on the data.

3. Model Comparison

In order to obtain the best model in estimating the probability of daily occurrences rainfall for every regions in West Java, we apply eight different models (Table 2).

Table 2: Eight different models

No	Model	Specification
1	Fixed Effect	$\eta_{it} = \beta_1 Time$
2	Spatially	$\eta_{it} = \beta_1 Time + v_i + \gamma_i$
3	Temporally	$\eta_{it} = \beta_1 Time + \omega_t + \phi_t + \varphi_t$
4	Spatial and temporally	$\eta_{it} = \beta_1 Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t$
5	Interaction (I)	$\eta_{it} = \beta_1 Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t + \delta_{it}(I)$
6	Interaction (II)	$\eta_{it} = \beta_1 Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t + \delta_{it}(II)$
7	Interaction (III)	$\eta_{it} = \beta_1 Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t + \delta_{it}(III)$
8	Interaction (IV)	$\eta_{it} = \beta_1 Time + v_i + \gamma_i + \omega_t + \phi_t + \varphi_t + \delta_{it}(IV)$

The following model selection criteria can be used. The most common Bayesian criterion is Deviance Information Criterion (DIC). DIC considers both fitness and complexity, defined as (Blangiardo et al., 2013):

$$DIC = \overline{D(\boldsymbol{\Phi})} + p_D$$

Equation 15

where $\overline{D(\boldsymbol{\Phi})}$ is the posterior mean of the deviance, defined as $\overline{D(\boldsymbol{\Phi})} = E[D(\boldsymbol{\Phi})]$, with $D(\boldsymbol{\Phi}) = -2\log(p(y|\boldsymbol{\Phi}))$, $p(\mathbf{y}|\boldsymbol{\Phi})$ the likelihood function, and p_D the effective number of parameters indicating model complexity. A lower DIC indicates a better fit. The second criterion is Marginal Predictive Likelihood (MPL) defined as (Urtasun, 2017):

$$\text{MPL} = \sum_{i=1}^{n}\sum_{t=1}^{T} \log(CPO_{it})$$

Equation 16

where CPO is the conditional predictive ordinate. The larger the MPL, the better the prediction. Other measures of predictability is Root Mean Square Error (RMSE) and the Pseudo Coefficient of Determination ($\widetilde{\mathrm{R}}^2$)

4. Result

The study is interested in estimating the underlying probability p_i of rainfall on a given calendar day $i = 1, \ldots, 365$, which is assumed to have spatial and temporal characteristics. We assume that the daily rainfall occurrence follows a Binomial distribution $Bin(n_{it}, p_{it})$, and associates p_{it} with spatial and temporal dependences. $iCAR$ prior is used to establish spatial autocorrelation and circular Random Walk of order 1 and seasonal prior for temporal autocorrelation. Circular model directly connects the end and the beginning of the temporal data. It is used to smooth change between the last week in December and the first week in January (Wang et al,. 2018). We then build the model and fit it in R-INLA which can be downloaded http://www.r-inla.org/ (e.g., model 8). Using INLA, the prior $iCAR$ can be defined using model="besag". For RW1 and seasonal models the priors are defined as model="rw1" and model="seasonal" respectively, where circular model of RW1 is defined using cyclic=TRUE. The exchangeable prior is defined by model="iid". Below we present the model selection by means the Bayesian (DIC, MPL) and frequentist (RMSEA, pseudo R^2) selection model criterions. Table 3 shows the results for the model selection criterion by means of MPL, DIC, RMSE, and pseudo R^2. All the criterion support the Model 8 provides a better pit model that the other models. It has largest MPL and R^2 and has smallest DIC, WAIC and RMSEA. Thus the best trade-off between model fit and complexity is offered by model 8. Figure 3 also supports that the model 8 is the best model where the predicted values is approximate to the observed values with the coefficient of correlation $\sqrt{0.886} = 0.940$ (Table 3).

```
prior.c=c(1,5*10^(-5))
hyper=list(theta=list(param=prior.c))
control <- list(
predictor = list(compute = TRUE,link=1),
results = list(return.marginals.random =
TRUE, return.marginals.predictor=TRUE),
compute = list(hyperpar=TRUE,
return.marginals=TRUE, dic=TRUE, mlik =
TRUE, cpo = TRUE, po = TRUE, waic= TRUE,
graph=TRUE, gdensity=TRUE,
openmp.strategy="huge"))

Model7<-y~-
1+f(ID,model="besag",graph=JABAR.GRAPH,cons
tr=FALSE,initial=1, hyper=hyper)+
 f(ID0,model="iid",param=c(1,0.00005))+
             IT3+f(IT,model="rw1",
constr=T,cyclic=TRUE, hyper=hyper)+
f(IT1,model="seasonal",
season.length=12,hyper=hyper,initial=365)+
             f(IT2,model="iid",
param=c(1,0.00005))+
             f(ID1,model="besag",
graph=JABAR.GRAPH, hyper=hyper, group=IT2,
control.group=list(model="rw1",cyclic=TRUE,
hyper=hyper))
ptm <- proc.time()
ModelRun7 <- inla(Model7,family="binomial",
Ntrials=n, data=DATA,
control.compute = control$compute,
control.predictor = control$predictor,
control.inla = list(int.strategy = "eb",
strategy = "gaussian"))
proc.time() - ptm
```

Table 3: Summary of the posterior mean the DIC as a measure of trade-off between model fit and complexity; additionally MPL, RMSEA, and R^2

Model	DIC	MPL	RMSEA	R^2
1	47805.180	-23902.660	2.551	0.002
2	37397.950	-18698.040	1.623	0.085
3	34512.990	-17219.690	1.312	0.402
4	33617.250	-16773.570	1.218	0.485
5	33618.770	-16773.940	1.218	0.485
6	33618.030	-16773.740	1.218	0.485
7	33617.670	-16773.600	1.218	0.485
8	30806.800	-15060.910	0.593	0.886

Table 4: Summary statistics of the posterior distribution of the fixed effect

Effect	Mean	Sd	Ratio = Mean/Sd
Trend Liner	-0.0027	0.0005	-5.4000

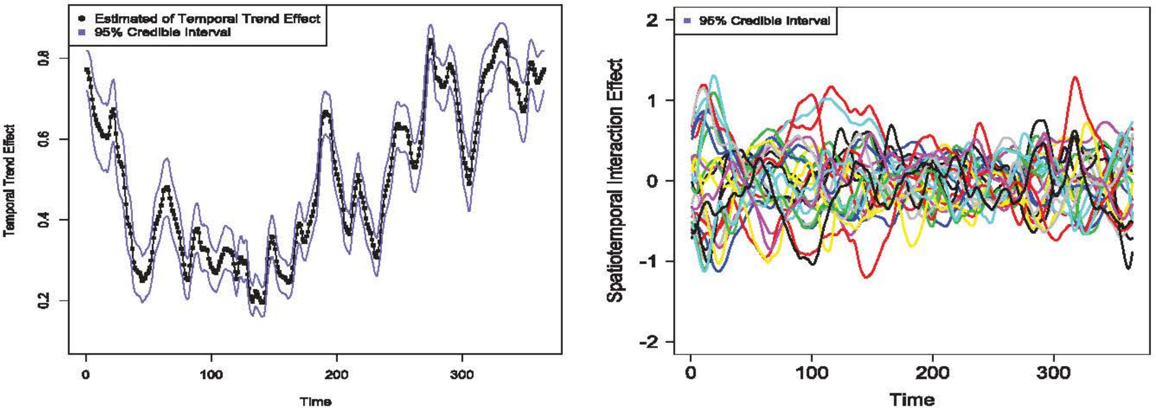
The fixed effect parameter of linear trend shows that there is a negative relationship between the probability of daily rainfall occurrence and the time in Table 3. It means in general the daily rainfall occurrences was decreased over times. However, the daily rainfall occurrence has spatial and temporal characteristics that maybe explain the underlying probability more clearly. Table 5 shows summary statistics of the posterior distribution of the random effects. Mean statistics indicates the mean of precision parameter ($1/variance$) of each random components. The ratio of mean on standard deviation (Sd) can be used to rank the most important random component in explaining the underlying probability of daily rainfall occurrence in West Java (Table 5).

Table 5: Summary statistics of the posterior distribution of the random effects

Components	Mean	Sd	Ratio = Mean/Sd
Spatially structured	16487.16	17964.115	0.918
Spatially Unstructured	20260.15	19577.183	1.035
Temporal Trend	131.19	17.42	7.531
Seasonal	51837.54	30603.65	1.694
Temporally Unstructured	26840.08	21343.104	1.258
Spatiotemporal Interaction	51.66	4.231	12.210

The most important component is spatiotemporal interaction type IV with the highest ratio value and the second component is circular temporal trend Random Walk of orders one. These two important components explain the underlying probability of daily rainfall occurrence in West Java. Here the spatially structured and unstructured seem less important because the spatial dependence was accommodated in the spatiotemporal interaction component. Figure 4 shows the effect of the three most important of random components. Figure (4a) presents the effect of the temporal components. This circular RW1 clearly present the underlying probability of daily rainfall in West Java. The highest probability of daily rainfall occurrence is around November and December. Those are the wet months in every year. The dry month between April-Jun. Figure (4b) presents the spatiotemporal interaction component. It indicates the underlying probability of daily rainfall occurrence in of a location will depend on the location of its neighbours. The two significant random components lead to the global and local pattern of underlying probability of daily rainfall occurrence in West Java.

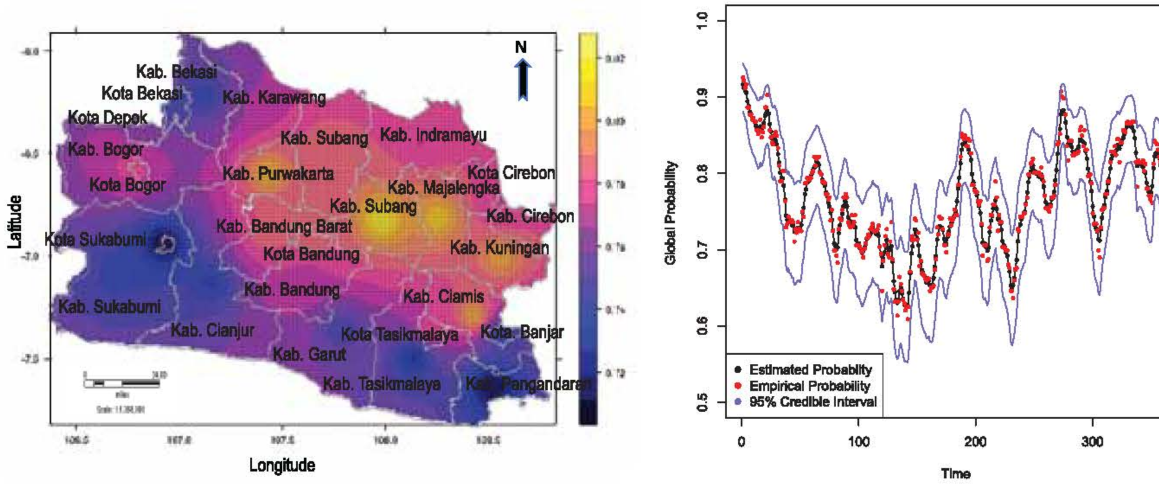
Figure 5 shows spatial and temporal distributions underlying probability of daily rainfall occurrence in West Java, 2010-2015. Figure (5a) presents isopleth map of the spatial distribution of estimated probability of daily rainfall occurrences which is almost similar to the empirical probability in Figure 1.



(a) Temporal Trend Component

(b) Interaction Component

Figure 4: The two most important random components (a) Temporal Trend and (b) Spatiotemporal Interaction components



(a) Spatial Distribution

(b) Temporal Distribution

Figure 5: Spatial and temporal distribution underlying probability of daily rainfall occurrence in West Java, 2010-2015

The isopleth map is used here instead of choropleth map in order to present more clear spatial pattern of the probability of daily rainfall occurrence. For the global model, we found that the Moran's Index is 0.205 (p-value=0.0425). It indicates that there is a significant spatial dependency between adjacent regions. Figure (5b) presents the temporal distribution. The red dots are the empirical probability and the black dots denote the estimated probability by means of logistic link function model. It seems that the typical chance of daily precipitation chance can reach 90%, and is relatively high in November up to February in every year. Those months are known as wet months while the dry months in April up to June. The underlying probability of daily rainfall occurrence is influence by temporal and spatiotemporal variation, it is important to present the underlying probability for 27 districts in West Java to explain the local underlying probability.

The local probability of daily rainfall occurrence by district is presented in Figure 5. Figure 6 displays the temporal empirical and estimated probability of daily rainfall occurrence for 30 districts in West Java, Indonesia. The empirical probability denotes by red dots and estimated by black dots. The significant effects of the circular temporal trend (RW1) and spatiotemporal interaction components produce a different temporal pattern of the probability of daily rainfall occurrence for every district. Some districts have daily precipitation chance over than 50% (e.g., Kab. Bogor and Kab. Majalegka) for every year. In general the highest probability in November up to February and the lowest probability in April until June.

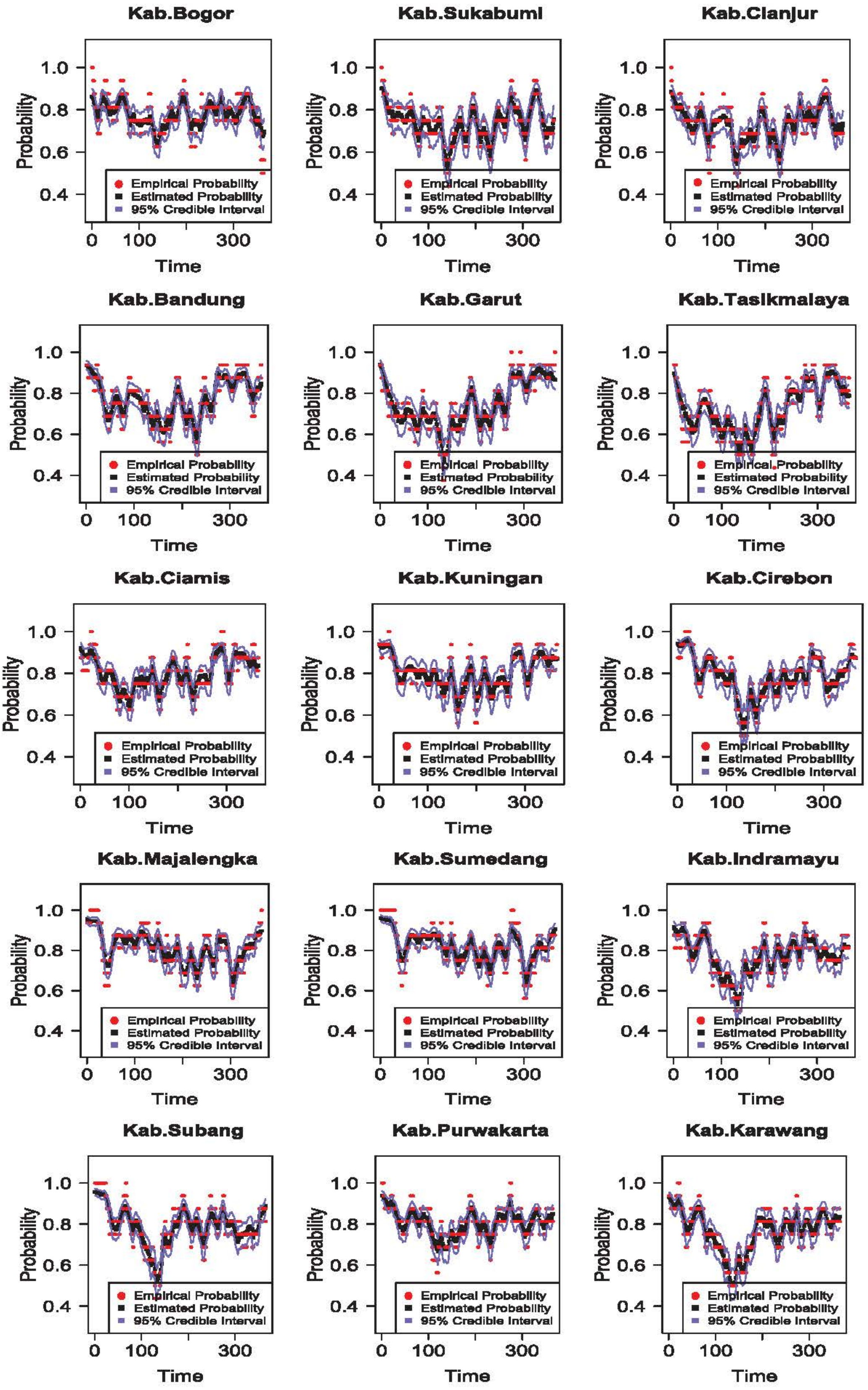


Figure 6: The Empirical and estimated probability of daily rainfall occurrence by districts (Continue next page)

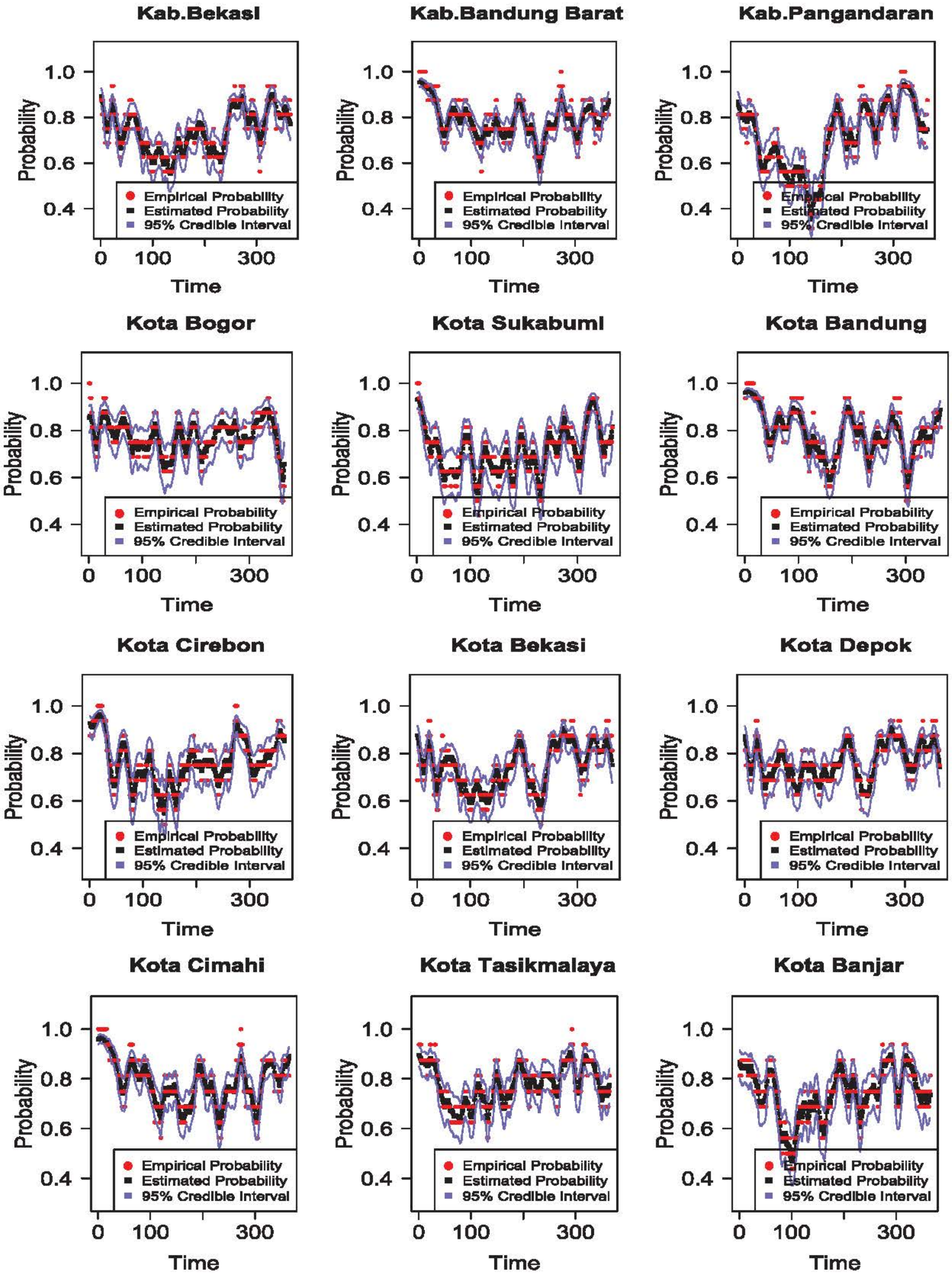


Figure 6: The Empirical and estimated probability of daily rainfall occurrence by districts

This information can play an important role in the society in every district. This information can be used to decide on the right time for agriculture and farming and also suggest a good time about travelling to these locations.

4. Discussion and Conclusions

Detailed models of the underlying probability of daily rainfall occurrence research are important tools in addressing many problems in the daily lives. They are often required to make a good prediction about the potential of flooding or drought.

This study presents an analysis of daily rainfall characteristics (frequency) in West Java, Indonesia from 2010 to 2015. The estimated properties of observed precipitation reveal many interesting features such as the spatial and temporal heterogeneity. This paper presents the novel idea of developing a model for predicting the underlying probability of daily rainfall occurrence using Bayesian spatiotemporal model. This paper provides a new approach of occurrences rainfall modelling using INLA. It is a powerful Bayesian approximation instead of MCMC in case the faster computing process (Rue and Martino, 2009). The spatiotemporal model which used in this analysis involves huge number of space-time parameters which are not possible to be calculated using the MCMC approach.

The model selection criterion (MPL, DIC, RMSE, Pseudo R^2) show that the model 8 is the best model for modelling the occurrences of rainfall in West Java, Indonesia. We find the three most important components in explaining the probability are fixed effect linear trend and random effects: circular temporal trend (RW1) and spatiotemporal components. Two latter components lead to the local probability of daily rainfall occurrence for every region. In addition, the interaction component types IV is the first component that has a significant contribution to explain the spatiotemporal pattern of the daily rainfall data. It indicates that the series pattern of the daily rainfall data of a region is related to the neighbouring regions. The temporal trend by means random walk of order 1 (RW1) is the second most important part of the model. The spatiotemporal analysis can also be used to identify the hot-spot of the regions which have high probability of rain every day. The empirical probability of daily rainfall occurrence in Figure 1 shows there is any cluster that means the data has spatial dependence characteristics. The southern regions of West Java have high probability to get rainfall every day.

The individual plots of the occurrence probability show the differences between occurrence probability with and without spatial dependence. The models which accommodate spatial dependence it is looked smoother. In general, the pattern shows that for every month there will be some rainy days. The daily rainfall data is obtained from CHIRPS data set with the consequences that the disaggregation of rainfall at the local scale would not be possible because the resolution of the CHIRPS data set. Such information on the disaggregation of rainfall could be used in a variety of tools to estimate the impacts of historic rainfall events or future rainfall scenarios in order to help decision makers respond to extreme climatic events (droughts, floods) (Kaptue et al., 2015).

Acknowledgement

This paper is funded by ALG Unpad contract: 2296/UN6.D/KS/2018 and RFU Unpad contract: 1732 d/UN6.RKT/LT/2018. The authors thank Rector Universitas Padjadjaran. The data contribution from LAPAN Bandung is also very much appreciated. The important and constructive recommendation to improve the quality of this paper is also gratefully acknowledged.

Reference

Besag, J., 1974, Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society*, Vol. 36(2), 192-236.

Besag, J., York, J. and Mollié, A., 1991, Bayesian Image Restoration, with Two Applications in Spatial Statistics. *Annals of the Institute of Statistical Mathematics*, Vol. 43(1), 1-20.

Bivand, R. S., Pebesma, E. and Rubio, V. G., 2013, *Applied Spatial Data Analysis with R*, Springer, New York.

Blangiardo, M., Cameletti, M., Baio, G. and Rue, H., 2013, Spatial and Spatio-Temporal Models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, Vol. 4, 33–49.

Blangiardo, M. and Cameletti, M., 2015, *Spatial and Spatio-temporal Bayesian Models with R – INLA*. John Wiley & Sons Ltd, United Kingdom

Chowdhury, K., Lockart, N., Willgoose, G., Kuczera, G., Kiem, A. and Manage, N. P., 2017, Development and Evaluation of a Stochastic Daily Rainfall Model with Long-Term Variability. *Hydrol. Earth Syst. Sci.*, Vol. 21, 6541- 6558.

Coe, R. and Stern, R., 1984, Modeling Fitting Analysis of Daily Rainfall Data. *J. R. Stat. Soc. A.*, Vol. 147, No.1, 1-34.

Jaya, I. G. N. M., Folmer, H., Ruchjana, B. N., Kristiani, F. and Andriyana, Y., 2017, Modeling of Infectious Disease: A Core Research Topic for The Next Hundred Year. *In J. Randall, & S. Peter, Regional Research Frontiers*. 681-701. US: Springer.

Jaya, I. G. N. M., Abdullah, A. S., Hermawan, E. and Ruchjana, B. N., 2016, Bayesian Spatial Modeling and Mapping of Dengue Fever: A Case Study of Dengue Fever in the City of Bandung, Indonesia. *International Journal of Applied Mathematics and Statistics*, Vol. 54(3), 94-103.

Jaya, I. G. N. M., Ruchjana, B. N., Tantular, B. Z., Andriyana, Y. and Ruchjana, B. N, 2018a, Simulation and Application of the Spatial Autoregressive Geographically Weighted Regression Model (SAR-GWR). *ARPN Journal of Engineering and Applied Sciences*, Vol. 13(1), 337-385.

Jaya, I. G. N. M., Toharudin, T. and Abdulla, A. S., 2018b, A Bayesian Spatial Autoregressive Model with k-NN Optimization For Modeling The Learning Outcome of the Junior High Schools in West Java, *Model Assisted Statistics and Applications*, Vol. 13(3), 207-219.

Jayasinghe, P., Bandara, N., Jayasinghe, A., Chemin, Y. and Mahanama, P., 2015, Assess the Level of Vulnerability to Climate Induced Disasters in Sri Lankan Coastal Areas: An Application of Open Source Geographic Information System. *International Journal of Geoinformatics*, Vol. 11(3), 53-58.

Johansson, B. and Chen, D., 2003, The Influence of Wind and Topography on Precipitation Distribution in Sweden: Statistical Analysis And Modelling. *International Journal of Climatology*, Vol. 23, 1523-1535.

Kaptue, A., Hanan, N., Prihodko, L. and Ramirez, J., 2015, Spatial and Temporal Characteristics of Rainfall in Africa: Summary Statistics for Temporal Downscaling. *Water Resources Research*, 51, 2668-2679.

Kitagawa, G., 1987, Non-Gaussian State-Space Modeling of Nonstationary Time Series with Discussion. *Journal of the American Statistical Association*, Vol. 82, No. 400, 1032–1063.

Knorr-Held, L., 2000, Bayesian Modelling of Inseparable Space - Time Variation in Disease Risk. *Statistics in Medecine*, Vol. 19, 15-30.

Kumar, R. and Bhardwaj, A., 2015, Probability Analysis of Return Period of Daily Maximum Rainfall in Annual Data Set of Ludhiana, Punjab. *Indian J. Agric. Res*, Vol. 49(2), 160-164.

Martins, T. G., 2014, October 25. About INLA. Retrieved August 1, 2017, from tgmstat.wordpress.com: https://tgmstat.wordpress.com/2013/10/16/latent-gaussian-models-inla/

Neethu, C. and Surendran, S., 2013, Review of Spatial Clustering Methods. *International Journal of Information Technology Infrastructure*, Vol. 2(3), 15-24.

Perera, H., Sonnadara, D. and Jayewardene, D., 2002, Forecasting the Occurrence of Rainfall in Selected Weather Stations in the Wet and Dry Zones of Sri Lanka. *Sri Lankan Journal of Physics*, Vol. 3, 39-52.

Roebber, P. J., and Bosart, L., 1996, The Complex Relationship between Forecast Skill and Forecast Value: A Real-World Analysis. *Weather and Forecasting*, Vol. 11, 544-559.

Rue, H., & Martino, S, 2009, Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Statist. Soc. B, 71(2), 319–392.

Singh, B., Rajpurohit, D., Vasishth, A. and Singh, J., 2012, Probability Analysis for Estimation of Annual One Day Maximum Rainfall of Jhalarapatan Area of Rajasthan. *Plant Archives*, Vol. 12(2), 1092-1100.

Smith, J. A., 1987, Statistical Modeling of Daily Rainfall Occurance. *Water Resources Research*, Vol. 23, 885-893.

Stauffer, R., Mayr, G., Messner, J., Umlauf, N. and Zeileis, A., 2017, Spatio-Temporal Precipitation Climatology over Complex Terrain using a Censored Additive Regression Model. *International Journal of Climatology*, Vol. 37, 3264–3275.

Thach, N. and Canh, P., 2013, Simulation of Flash-Muddy Flash Flooding and Inundation of Western Tamdao Mountain Region, Vinh Phuc Province, Viet Nam using Hydro-geomorphology. *International Journal of Geoinformatics*, Vol. 9(2), 39-47.

Urtasun, A. A., 2017, Hierarchical and Spline-Based Models in Space-Time Disease Mapping. Disertation, Universidad Publica de Navarra

Wang, X., Yue, Y. R. and Faraway, J., 2018, Bayesian Regression Modeling with INLA. Boca Raton: Taylor & Francis Group, LLC.

Wilson, P. and Toumi, R., 2005, A Fundamental Probability Distribution for Heavy Rainfall. *Geophysical Research Letters*, Vol. 32, 1-4.