Parameter Estimation Using INLA for Disease Mapping of Leptospirosis in Bantul Indonesia

R Artiono
Mathematics Department, Universitas Negeri Surabaya, Surabaya, Indonesia
e-mail: rudiantoartiono@unesa.ac.id

Abstract. Agriculture sector is recognized as one important factor for an increased risk of leptospirosis in Bantul, Indonesia. From January to July 2010, there were 70 patients of leptospirosis with 43 patients engaged in agriculture. The aim of this study was to determine the impact of agricultural sector on the disease mapping of leptospirosis in district level. A structured additive regression model was used for statistical analysis associated with agricultural area and agricultural livelihood. In addition, a random effect that captures the spatial heterogeneity in the study region was also included in this study. Model estimation and predictive inference was carried out through the implementation of computer code in the R-INLA package, which makes use of the Integrated Nested Laplace Approximation.

1. Introduction
Leptospirosis is a bacterial disease of worldwide importance. This disease can spread sporadically not only in the human population but also in the domestic animal population. The disease is transmitted by rats, which act as a reservoir for Leptospira bacteria. The bacteria can breed well in the rat body but they do not infect the rat itself. Thus, the favourable environments for rats are associated with the increasing of the risk of leptospirosis.

Generally, incidence is associated with lack of access to clean water supplies [1,2]. A contaminated water provides favourable conditions for the growth of bacteria. It is also well-known that high rainfall can increase the number of human cases of leptospirosis [3-5]. The survival rate of bacteria is known to change seasonally, mortality being lower in rainy season and higher in dry season [6].

The cases can be found not only in tropical areas but also in the sub tropics. The World Health Organisation (WHO) reported that in the period between January 1970 and October 2008, there were 12,033 leptospirosis cases in United Kingdom, France, Italy, and Spain [7]. Another investigation reported a total of 2,694 cases of Leptospirosis in Germany 1962-2003 and a total of 2,553 cases of leptospirosis in Netherland 1925-2008 [8,9]. Moreover, in tropical areas, some countries reported individual years of high incidence of leptospirosis. (e.g. Thailand and Philippines). The Thailand Department of Disease Control (DDC) reported a total of 14,285 cases of leptospirosis in the year 2000 and the National Epidemiology Centre of Philippines also reported about 2,471 cases of leptospirosis in the year 2012 [10,11].

Indonesia, as a country that lies in the tropical area, also has a significant number of cases. Data obtained from the Health Ministry of Indonesia show that Leptospirosis in humans fluctuated from 2005 to 2010. In the years 2005 to 2007, the number of cases increased dramatically from 115 patients to be 664 patients. However, in 2008 and 2009, it dropped to 426 patients and 335 patients respectively. The number of cases increased again in 2010 to be 409 patients. Although, Indonesia has relatively few cases...
of leptospirosis compared to Thailand and Philippines, it does not mean that Indonesia has low potential risk to the leptospirosis disease; it is because the disease is well-known that there are undiagnosed cases, misdiagnosed cases, and unreported cases of leptospirosis [12,13].

One area in Indonesia that has a lot of cases of leptospirosis is Bantul, small area in the Yogyakarta province. It is one county with an area of 508.85 km$^2$. The latitude is -7.88806 and the longitude is 110.32889. Based on the population census done by the Central Bureau of Statistics of Bantul in 2010, Bantul has population as much as 910,572 with the main livelihood is in agriculture. The Health Department of Bantul have reported that in 2009, it was found 9 cases of leptospirosis with 1 person died because of this disease. The number of cases had increased in 2010 became 116 cases with 19 people died and grew up in 2011 became 154 cases with 12 people died. In the year 2010 and 2011, the local government of Bantul stated that Bantul is endemic of leptospirosis. The increasing of the leptospirosis cases occurs more often in the rainy season. It was found that in the period of February-April is the peak season of the spread of leptospirosis disease.

In this study, we used data of 70 patients with leptospirosis in 17 districts from January to July 2010. The data were about the number of cases, the percentage of agricultural land used and the percentage of livelihoods associated with agriculture.

In epidemiology, the most important thing that we need to concern is about how to make inference in order to analyse the spread of disease from some previous data. Unfortunately, it is not easy to get complete data in public health area. There are some aspects which relate to each other. The increasing number of cases is not only depending on the disease itself. It could be some aspects associate with it (i.e. environment, job, life style, etc.). However, it is still possible to construct a stochastic model which can be used as a starting point to inference. Particularly, models attempt to describe the mechanism by which the observed data are generated. Inference then proceeds by attempting to estimate the parameters of the model. One method that can be used to estimate the parameter is Bayesian. This method comes from the extension of the likelihood paradigm. In likelihood paradigm, it is only known that population are random variable. Meanwhile in Bayesian, it is not only population that are random variable but parameters are also random variable. Thus, the parameter also have distribution and we called it as prior distribution. In order to estimate the parameter, we combine likelihood model (data) and prior distribution such that posterior distribution can be determined. This process can be continuously updated. It means that posterior distribution could be prior distribution for the next estimation.

The primary goal of this study was to determine the impact of agricultural sector on the disease mapping of leptospirosis in district level with Integrated Nested Laplace Approximation (INLA). This approximation was used to find posterior probability and Standard Morbidity Rate. We hypothesized that agricultural area and agricultural livelihood have big influence to the spread of leptospirosis in Bantul.

Calculation and plotting data in this study were conducted entirely by using R. This software also provides INLA package so that in order to find posterior marginal distribution with INLA method could be determined easily. Through shape-file in one area, we could make plot for the posterior probability
and the standard morbidity rate. ArcGIS software was used for showing some data related to the Geographic Information System. This software could be used to obtain the shape-file which we need to plotting data in R.

2. Latent Gaussian Model and Hierarchical Structure
Model which used in Bayesian disease mapping is regression models [14-17]. The correlation between dependent variable and independent variable can be determined through this model. Furthermore, regression model that focused on the spatial elements is the spatial models. It is a subclass of structure additive regression models. This model allows us to be able to include some covariates either directly related to spatial or not. This model is more flexible to use. One of the models that exist in the structure additive regression models is Latent Gaussian Models.

Suppose $y_i$ is the number of patients in each district area in Bantul which is modelled as

$$y_i \sim \text{Poisson}(\lambda_i);$$

with link function $\eta_i = \log \log (\rho_i)$, define latent Gaussian field as $x = (\eta, \alpha, f, \beta)$. Since $\eta, \alpha, f$ and $\beta$ are random variable then it should have parameter. Suppose $f(i)$ is parameter for $y_i$ and $\psi$ is parameter for $x$ then we can build a hierarchical structure for this.

Stage 1
Latent Gaussian Model

$$\eta_i = \log \log (\rho_i) = \alpha + v_i + v_i + \beta_1 Z_{1i} + \beta_2 Z_{2i}$$

where:

- $\lambda_i$ is the Mean
- $E_i$ is the Expected Number of Cases
- $\alpha$ is Intercept Term
- $v_i = f_1(i)$ is Spatially Structured Residual
- $v_i = f_2(i)$ is Spatially Unstructured Residual
- $\beta_1 Z_{1i}$ is linear effect of covariate Agricultural Area
- $\beta_2 Z_{2i}$ is linear effect of covariate Agricultural Livelihood

Stage 2
Considering spatial structure, it gave two distributions as follows:

1. $v_i = f_1(i)$ Spatially Structured Residual
   Modeled using an intrinsic conditional autoregressive structure
   $$v_i \mid v_{j \neq i} \sim \text{Normal}(m_i, s_i^2)$$
   $$m_i = \frac{\sum_{j \in N(i)} v_j}{#N(i)}$$
   $$s_i^2 = \frac{\sigma^2_{\alpha}}{#N(i)}$$
   Where: $#N(i)$ is number of areas which share boundaries

2. $v_i = f_2(i)$ Spatially Unstructured Residual
   Modeled using an exchangeable prior
   $$v_i \sim \text{Normal}(0, \sigma^2_v)$$

Stage 3
$$\psi = \{\sigma^2_{\alpha}, \sigma^2_v\}$$
Furthermore, informative prior are specified by INLA’s default, as follows:

$$\sigma^2_{\alpha} \sim \text{Gamma}^{-1}(1,0.0005)$$
$$\sigma^2_v \sim \text{Gamma}^{-1}(1,0.0005)$$

The framework can be seen in the diagram below.
Figure 2. Framework of Hierarchical Structure

It represents spatial structure in each area satisfy Gaussian Markov Random Field (Conditional Independence).

3. Parameter Estimation

The fixed parameters effect \((\alpha, \beta_1, \beta_2)\) estimated by INLA are presented in Table 1. If it is exponentiated, they can be interpreted as relative risks: an increase of one unit in the agricultural land and in the agricultural livelihood is associated respectively with an increase of around 3% = \(\exp(0.036)\) and around 0.3% = \(\exp(0.0035)\) in the spread of disease.

| Parameter | Mean               | Sd      | 0.025quant  | 0.5quant   | 0.975quant  |
|-----------|--------------------|---------|-------------|------------|-------------|
| \(\alpha\) | -2.421907361       | 2.18184235 | -7.33178841 | -2.258992729 | 1.51788256  |
| \(\beta_1\) | 0.035703917        | 0.04279122 | -0.04250809 | 0.032720051  | 0.13180982  |
| \(\beta_2\) | 0.003532001        | 0.002851445 | -0.05580304 | 0.004001843  | 0.0601778   |

Agricultural land and agricultural livelihood just give a small effect to the spread of the leptospirosis diseases. Both parameters only give effect less than 10% in every increasing of one unit.

4. Posterior Probability

Furthermore, by using INLA, we can produce the posterior probability that represent the probability of finding cases of leptospirosis disease in one area after some factors (i.e. agricultural area and agricultural livelihood) are considered. The proportion of people which probably will be infected of leptospirosis can be obtained. Agricultural area represents the percentage of land area which is used for agricultural purposes, while agricultural livelihood is the percentage of population number which works in agriculture sector. The interpretation of the probability value is divided into three levels:

a. Area with the probability value between 0.6-1, it means that we are sure in those area will have a big probability to the spread of the Leptospirosis disease or in the other word we will find a lot of new cases of leptospirosis disease on those area or we are going to find a new patient of leptospirosis disease easily on those area. Area with this probability value is defined as areas with high risk to the spread of the Leptospirosis disease.

b. Area with the probability value between 0.3-0.6, it means that we are sure those area will have probability to the spread of the Leptospirosis disease but it is not too dangerous or in the other word we will find new cases of leptospirosis disease on those area but it is not as many as area with the probability value 0.6-1. Area with this probability value is defined as areas with medium risk to the spread of the Leptospirosis disease.

c. Area with the probability value between 0-0.3, it means that we sure those area will have a small probability to the spread of the Leptospirosis disease or in the other word we will not find new cases of leptospirosis disease on those area or it is not easy to find a new patient of leptospirosis disease on those area. Area with this probability value is defined as areas with low risk to the spread of the Leptospirosis disease.
The posterior probability, which was obtained from arcGIS and INLA, can be seen in Figure 3, 4, 5, and 6.

Figure 3 represents the probability of finding case of leptospirosis disease in each district of Bantul Area without considering of Agricultural Area and Agricultural Livelihood. Figure 4 represents the probability of finding cases of leptospirosis disease in each district of Bantul Area by considering of Agricultural Area. Figure 5 represents the probability of finding cases of leptospirosis disease in each district of Bantul Area by considering of Agricultural Livelihood. Figure 6 represents the probability of finding cases of leptospirosis disease in each district of Bantul Area by considering the agricultural area and livelihood. The conclusions which can be derived from figure 3-6 are as follows:

a. Areas such as Sedayu, Banguntapan, Pleret, Imogiri, Jetis, Bambanglipuro, Sandan and Kretek have a high risk for the spread of the leptospirosis disease. Although some areas such as Pleret, Piyungan, and Kretek did not reveal any leptospirosis patients, but this area includes in the area with high risk against the spread of the disease leptospirosis.

b. Some areas with a relatively large number of patients as Sewon and Bantul did not have a high risk to the spread of leptospirosis disease.

c. There is no significant change in the posterior probability when agricultural area is added to the model as covariates. All area has same risk to the spread of leptospirosis disease. It means that agricultural area cannot be a major influence in the spread of the leptospirosis disease.

d. There is also found no significant change in the posterior probability when agricultural livelihood is added to the model as covariates.

e. The significant change appears when both of them are combined. Posterior probability in Bambanglipuro area is decrease from the range 0.6-1 to the range 0.3-0.6 and Dlinggo area is increase from the range 0-0.3 to the range 0.3-0.6. It means that Bambanglipuro area possible to
decrease the level of risk of the leptospirosis disease while conversely Dlingo area increase the risk of the spread of Leptospirosis disease.

f. Areas such as Kasihan, Sewon, Pajangan, Bantul, Pandak, and Pundong remain secure against the spread of leptospirosis disease because it has a posterior probability in the range 0-0.3. These areas are classified as areas with the low risk for the spread of disease.

5. Standard Morbidity Rate

Standard Morbidity Rate is the proportion of new cases of disease that found in one area in a given time. By run the INLA program, SMR in Bantul area can be seen in Figure 7, 8, 9, and 10.

Figure 7 represents the Standard Morbidity Rate of leptospirosis disease in each district of Bantul Area without the considering of Agricultural Area and Agricultural Livelihood. Figure 8 represents the Standard Morbidity Rate of leptospirosis disease in each district of Bantul Area by considering Agricultural Area. Figure 9 represents the Standard Morbidity Rate of leptospirosis disease in each district of Bantul Area by considering Agricultural Livelihood. Figure 10 represents the Standard Morbidity Rate of leptospirosis disease in each district of Bantul Area by considering Agricultural Area and Agricultural Livelihood.

The conclusions which can be derived from figure 6-9 are as follows:

a. If it does not involve the covariate, then Banguntapan has the highest SMR in the interval 9-15; Piyungan, Imogiri, and Sanden have SMR in the interval 6-9; Sedayu, Kasihan, Sewon, Pleret, Dlingo, Jetis, Bantul, Pajangan, Pandak, Bambanglipuro, Pundong, and Kretek have SMR in the interval 3-6, and Srandakan has the lowest SMR in the interval 0-3.

b. There is significant change that occurred in Banguntapan when agricultural area and/or agricultural livelihood was considered as covariate in the model. SMR of this area has decreased to the range 0-3.

c. The number of SMR in some areas such as Sedayu and Kretek increase when agricultural area was included in the models as covariates. While other areas did not change significantly.
When the agricultural livelihood is included in the model then area such as Imogiri and Jetis have the decreasing of number from 6-9 to 0-3 and from 3-6 to 0-3.5 respectively. There is no significant change when two covariates are included in the model. All of the area has the same SMR when agricultural livelihood was considered as covariate.

6. Conclusions
By using INLA, the parameter effect of agricultural land and agricultural livelihood can be obtained. The effect of agricultural land is relatively bigger than agricultural livelihood. However, both of them have quite small effect to the spread of disease. They only give an effect to the spread of the disease less than 10%. INLA also can be used to explore posterior probability and standard morbidity rate in Bantul area. Some area has high risk of new cases of leptospirosis.

References
[1] Walker M, Wilcox B A and Wong M 2008 Waterborne zoonoses and changes in hydrologic response due to watershed development, Coastal Watershed Management
[2] Messier V, Levesque B, Proulx, J F, Rochette L, Serhir B, Couillard M, Ward B J, Libman M D, Dewailly E and Dery S 2012 Seroprevalence of Seven Zoonotic Infections in Nunavik, Quebec (Canada) Zoonoses and Public Health
[3] Kupek E, de Sousa Santos Faversani M C and de Souza Philippi J M 2000 The relationship between rainfall and human leptospirosis in Florianopolis, Brazil, 1991-1996, The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases
[4] Ward M P 2002 Seasonality of canine leptospirosis in the United States and Canada and its association with rainfall, Preventive Veterinary Medicine
[5] Davis S, Calvet E and Leirs H 2005 Fluctuating rodent populations and risk to humans from rodent-borne zoonoses, Vector-borne and Zoonotic Diseases 5 pp 305-314
[6] Holt J, Davis S and Leirs H 2006 A model of Leptospirosis infection in an African rodent to determine risk to humans: Seasonal fluctuations and the impact of rodent control, Acta Tropica
[7] World Health Organization 2010 Report of the First Meeting of the Leptospirosis Burden Epidemiology Reference Group
[8] Jansen A, Schneberg I, Frank C, Alpers K, Schneider T and Stark K 2005 Leptospirosis in Germany 19622003 Emerg Infect Dis. 11(7) 10481054.
[9] Goris M, Boer K, Duarte T, Klien S, and Hartskeerl R, 2013 Human Leptospirosis Trends, the Netherlands, 19252008 Emerging Infectious Disease 19(3)
[10] Wuthiekanun V, Chierakul W, Limmathurotsakul D, Smythe L D, Symonds M L, Dohnt M F, Slack A T, Limpaiboon R, Suputtamongkol Y, White N J, Day N P, Peacock S J 2007 Optimization of Culture of Leptospira from Humans with Leptospirosis, J Clin Microbiol 45(4) 13631365
[11] Klutzfree 2012 Leptospirosis disease in the Philippines. http://news.wikinut.com/Leptospirosis-disease-in-the-Philippines/inju6u2f
[12] Cachay E R and Vinetz J M 2005 A global research agenda for leptospirosis J Postgrad Med 51 p 174
[13] Levett P N 2001 Leptospirosis Clin Microbiol Rev 14(2) pp 296-326
[14] Blangiardo M, Camalette M, Baio G and Rue H 2013 Spatial and spatiotemporal models with R-INLA Spat Spatiotemporal Epidemiol 4 pp 33-49
[15] Lawson B A 2009 Bayesian Disease mapping: Hierarchical Modelling in Spatial Epidemiology, Chapman & Hall/CRC Interdisciplinary Statistics Series
[16] Rue H, Martino S and Chopin N 2009 Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71 319392
[17] Martins G T, Simpson D, Lindgren F, and Rue H, 2013 Bayesian computing with INLA: new features, arXiv:1210.0333 [stat.CO]