Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Association of Population Density and Distance to the City with the Risks of COVID-19: A Bayesian Spatial Analysis

To cite this article: M A Tiro *et al* 2021 *J. Phys.: Conf. Ser.* **2123** 012001

Association of Population Density and Distance to the City with the Risks of COVID-19: A Bayesian Spatial Analysis

M A Tiro[1], A Aswi[1 *], and Z Rais[1]

[1] Statistics Department, Universitas Negeri Makassar, Indonesia

[*]Email: aswi@unm.ac.id

Abstract. The outbreak of Coronavirus disease-2019 (Covid-19) poses a severe threat around the world. Although several studies of modelling Covid-19 cases have been done, there appears to have been limited research into modelling Covid-19 using Bayesian hierarchical spatial models. This study aims to examine the most suitable Bayesian spatial CAR Leroux models in modelling the number of confirmed Covid-19 cases without and with covariates namely distance to the capital city and population density. Data on the number of confirmed positive cases of Covid-19 (March 20, 2020 - August 30, 2021) in 15 sub-districts in Makassar City, the number of populations, population density, and distance to the city are used. The best model selection is based on several criteria, namely Deviance Information Criteria (DIC), Watanabe Akaike Information Criteria (WAIC), residuals from Moran's *I* Modification (MMI), and the 95% credible interval does not contain zero. The results showed that the best model in modelling Covid-19 is spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) model with the incorporation of distance to the capital city. It is found that there was a negative correlation between the distance to the capital city and Covid-19 risk, but the association between population density and the relative risk of Covid-19 was not statistically significant. Ujung Pandang district and Sangkarrang Island have the highest and the lowest relative risk respectively.

Keywords: Bayesian Spatial, Conditional Autoregressive priors, Population density, Distance

1. Introduction

The first case of coronavirus disease 2019 (Covid-19) was reported in Indonesia on March 2, 2020. It is reported that Indonesia has approximately 4,079,267 confirmed cases of Covid-19 from March 2, 2020 to August 30, 2021 with about 3,743,716 recovered and 132,491 deaths (https://infocorona.makassar.go.id/) [1]. South Sulawesi province has the highest number of confirmed Covid-19 cases outside Java Island [2]. Among 24 districts/cities in South Sulawesi Province, Makassar city has the highest number of Covid-19. It is reported that there were approximately 47,293 confirmed cases with at least 44,243 recovered and 954 deaths (19 March 2020 - 30 August 2021).

Although a number of studies of modelling Covid-19 cases have been done, there appears to have been limited research into modelling Covid-19 using Bayesian hierarchical spatial models. Hierarchical models include the random effect model, and generalized linear (mixed) model (GLM/GLMM) [3]. One of the advantages of Bayesian hierarchical models is that they enable for sharing of information across adjacent responses and allow the inclusion of covariates in the model [4]. Recently, a Bayesian spatial conditional autoregressive (CAR) Besag, York & Mollié (BYM) model where y_i counts in area i are modelled using Poisson and Negative Binomial have been compared to examine the relationship between the number of positive Covid-19 and socioeconomic factors in New York City [5]. They found that there is a statistically significant correlation between the number of Covid-19 cases and population density, race, household income and children under 18 years old.

Bayesian hierarchical spatial BYM models using Poisson for count data y_i was also used to analyze the relationship between positive covid-19 and socioeconomic and health [6]. They found that the proportion of populations older than 65 years, the proportion of the population with heart disease, and housing density were positively associated with Covid-19 cases. They also conclude that the proportion of black/African American populations is associated with the number of positive Covid-19 cases. The Bayesian hierarchical spatial using re-parametrization of the BYM models was also used to investigate the relationship between Covid-19 mortality and air pollution (NO2 and PM2·5) in England [7].

The association between the distance from the virus epicenter, population density, the ratio of the population older than 65 years, and the incidence of Covid-19 in Iranian provinces have been evaluated by using linear regression analysis [8]. The results show that the incidence of Covid-19 was strongly negatively associated with the distance from the virus epicenter. They also found that the ratio of the population older than 65 years was positively associated with the incidence of Covid-19 incidence. However, the association between population density and the Covid-19 incidence was not statistically significant [8].

The performance of different Bayesian spatial models such as conditional autoregressive BYM [9], CAR Leroux [10], CAR localised has been investigated and found that overall, the Leroux model performed better than the CAR BYM model [11]. The association of the population density, the distance from the virus epicenter, and the Covid-19 incidence have been done, but they used linear regression [8]. To our knowledge, the Bayesian hierarchical spatial CAR Leroux model has not been investigated in modeling Covid-19 risk. The association of the population density, the distance to the capital city, and the Covid-19 risk have not been explored yet. This paper aims to examine the most suitable Bayesian spatial CAR Leroux models in modelling the number of confirmed Cocid-19 cases without and with covariates (distance to the capital city and population density).

2. Methods

2.1 Study Area

Makassar city is situated between 119^0 24'17'38" East Longitude and 5^0 8'6'19" South Latitude. It covers an area of 175.77 km square ranging from 1.54 km square (Sangkarrang Island) to 48.22 km square (Biringkanaya). The city has 15 districts namely Biringkanaya, Bontoala, Makassar, Mamajang, Manggala, Mariso, Panakkukang, Rappocini, Sangkarrang Island, Tallo, Tamalanrea, Tamalate, Ujung Pandang, Ujung Tanah, and Wajo districts. Makassar has approximately 1,423,877 population ranging from 14,125 (Sangkarrang Island) to 209,048 (Biringkanaya) in 2020. Population density in 2020 attained 228.231 people per km square. Makassar district has the highest population density (32566) followed by Mariso (31553), while Tamalanrea has the lowest population density (3240) followed by Biringkanaya (4335) [12]. Ujung Pandang district is the capital city of Makassar and the administrative centre of Makassar city. Biringkanaya has the longest distance (21 km) to the capital (Ujung Pandang), followed by Sangkarrang island (20 km). The shortest distance from Ujung Pandang is Makassar district (1.8 km) followed by Bontoala (1.9 km) [12].

2.2 Covid-19 data

The number of confirmed cases of Covid-19 data from 19 March 2020 to 30 August 2021 for each district was obtained from the official website of Makassar Health Office (https://infocorona.makassar.go.id/) [1]. Population data in each district were gathered from the Badan Pusat Statistik (BPS) [12]. Population density and the distance to the capital (Ujung Pandang) which were also used in this study were gathered from the Badan Pusat Statistik (BPS) [12].

2.3. Models

One measure of disease risk is the Standardised Incidence Ratio (SIR) which is defined as the ratio of the observed counts to the expected counts. However, the SIR may be inadequate when the diseases are rare and the population is small. Bayesian models are suggested to estimate disease risk because they enable the incorporation of information from neighboring areas and covariates in the model.

In this study, the Bayesian spatial CAR Leroux [10] was used in estimating the Covid-19 risk in Makassar, Indonesia as well as quantifying the risk associated between Covid-19 and covariates namely population density and the distance of each district to the capital city. Population density is defined as the number of populations divided by the area (the number of populations per square kilometer). Bayesian spatial CAR Leroux model consists of one component, namely spatial random effect (u_i) which allows with different spatial autocorrelation (ρ) between zero and one.

The confirmed Covid-19 counts (y_i) were modelled using the Poisson distribution with mean is the expected count (E_i) times the relative risk of Covid-19 in the *i*th area (θ_i). A logarithmic transformation enables a linear, additive model regression model along with a spatial random effect (u_i). All model combinations were analysed using the CARBayes package version 5.2.3 [13] in the software R version 3.6.1 [14]. The model is explained as follows.

$$y_i \sim \text{Poisson}(E_i \theta_i)$$
$$\log(\theta_i) = \alpha + \beta X + u_i$$

α and $\beta = (\beta 1, \beta 2, \ldots, \beta p)$ are the overall levels of relative risk and the coefficients of the covariates respectively. θ_i is the relative risk of Covid-19 in the *i*th area and u_i, the spatial structure random effect, is defined by a conditional autoregressive (CAR) prior as follows.

$$\left(u_i \middle| \boldsymbol{u}_j, i \neq j, \tau_u^2\right) \sim N\left(\frac{\rho \sum_j u_j w_{ij}}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\tau_u^2}{\rho \sum_j w_{ij} + 1 - \rho}\right)$$

The spatial adjacency matrix w_{ij} that quantify spatial between areas i and j is defined using the binary neighbourhood matrix [15, 16]:

$$w_{ij} = \begin{cases} 1 & \text{if areas } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases}$$

where $w_{ii} = 0$ and $w_{ij} = w_{ji}$.

Modified Moran's *I* (MMI) [17, 18] and Moran's *I* statistics [19] for the observed data as well as for the residuals from the model were computed to quantify the spatial autocorrelation.

A sensitivity analysis was also performed to investigate the effect of the prior on the estimation of posterior quantities. Four different priors on the precision terms were used namely, Inverse-Gamma (1, 0.01), the default hyperprior in CARBayes, Inverse-Gamma (0.5, 0.05), Inverse-Gamma (0.1, 0.01), and Inverse-Gamma (1, 0.1).

Model formulations and combinations of covariates were compared using the 95% posterior credible interval (considered substantive when the interval does not contain zero), and the goodness of fit measures, Deviance Information Criterion (DIC) [20], Watanabe Akaike Information Criterion (WAIC) [21] and Modified Moran's *I* (MMI) [17, 18] for the residual from the model.

3. Results

The first three highest number of confirmed Covid-19 cases in Makassar city is Rappocini district (6,602), Biringkanaya district (6,166), and Tamalate district (5,497), while the first three lowest number of confirmed Covid-19 cases is Sangkarrang Island (38), Ujung Tanah (639), and Wajo (1106). The first three highest population density is Makassar district (32566), Mariso (31553), and Bontoala district

(26189), while the first three lowest population density is Tamalanrea (3240), Biringkanaya (4335), and Manggala (6078). The value of Moran's I for observed data is 0.46 (p-value = 0.0009) which suggests that there is a positive spatial autocorrelation in observed data. The value of MMI is relatively similar (0.42).

The results of Bayesian Spatial CAR models for all four different hyperpriors (τ_u^2) without and with covariates for confirmed Covid-19 cases from March 20, 2020, to August 30, 2021, with four distinct options for hyperprior (τ_u^2) are given in Table 1 and demonstrate insensitivity to the choice of hyperprior.

Table 1. Bayesian spatial CAR leroux model without and with covariates for all 4 distinct hyperpriors for confirmed covid-19 cases from March 20, 2020, to August 30, 2021

Hyperprior (τ_u^2)	**Models**	**DIC**	**WAIC**	**Residual using MMI**	**Posterior Quantities for Covariates**	
					2.5%	97.5%
Inverse-Gamma(1, 0.01)	Without Covariates	176.96	176.92	0.11		
	Distance	177.04	177.55	0.11	-0.97	-0.25
	Pop Density	177.60	179.63	-0.38	-0.51	0.69
	Distance+	177.021	178.59	0.47	-1.09	-0.48
	Pop Density				-0.71	-0.01
Inverse-Gamma(0.5, 0.05)	Without Covariates	177.00	177.42	-0.36		
	Distance	**176.50**	**175.93**	**-0.15**	**-0.91**	**-0.12**
	Pop Density	177.63	178.86	-0.72	-0.42	-0.49
	Distance+	176.78	178.48	-0.70	-1.32	-0.22
	Pop Density				-0.67	-0.02
Inverse-Gamma(0.1, 0.01)	Without Covariates	**176.17**	**175.89**	-0.49		
	Distance	177.09	177.44	-0.13	-0.96	-0.16
	Pop Density	177.32	178.31	-0.04	-0.51	0.3
	Distance+	177.21	178.88	-0.19	-1.10	-0.34
	Pop Density				-0.65	-0.11
Inverse-Gamma(1, 0.1)	Without Covariates	176.87	177.04	-0.26		
	Distance	177.14	178.09	-0.34	-0.93	-0.15
	Pop Density	177.26	178.13	0.01	-0.42	0.33
	Distance+	177.42	178.69	-0.14	-1.18	-0.23
	Pop Density				-0.88	0.06

Based on Table 1, it can be seen that Bayesian Spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) incorporate with covariate distance to the capital city has the lowest DIC (176.50) and WAIC (175.93). The DIC (176.17), as well as WAIC (175.89), are indistinguishable from that for the Bayesian Spatial CAR Leroux with hyperprior Inverse-Gamma (0.1, 0.01) model without covariates. The MMI for residual using Spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) is closer to zero (- 0.15) than Spatial CAR Leroux with hyperprior Inverse-Gamma (0.1, 0.01) model (-0.49).

All four models indicated that distance to the capital city was considered significant as the 95% posterior credible interval for the coefficient does not include zero. However, three models found that population density was not considered significant as the 95% credible interval contain zero. Only one model indicated that the population density was significant (model with hyperprior Inverse-Gamma (0.5, 0.05). Furthermore, Bayesian Spatial CAR Leroux with hyperprior Inverse-Gamma (1, 0.1) with the inclusion of both the distance to the capital city and population density was not significant. While using the other three hyperpriors suggested that the inclusion of both the distance to the capital city and population density were significant.

The number of confirmed Covid-19 cases, population density, distance to the capital city, and the relative risk (RR) value for each district based on the best model, namely Bayesian Spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) incorporate with covariate distance to the capital city are given in Table 2.

Table 2. The number of confirmed covid-19 cases, population density, distance to the capital city, and the relative risk (RR) values for each district based on the best model

ID	Districts	Number of Covid-19 Cases	Population Density	Distance to the Capital City	RR
8	Ujung Pandang	1323	9325	0	1.71
5	Rappocini	6602	15665	7.5	1.45
14	Tamalanrea	4507	3240	13	1.39
11	Wajo	1106	15061	2.9	1.17
9	Panakukkang	5064	8187	8	1.15
2	Manggala	5042	6078	11	1.09
1	Mamajang	1884	24911	2.9	1.07
6	Tamalate	5497	8947	6.2	0.97
15	Biringkanya	6166	4335	21	0.94
3	Mariso	1648	31553	2.9	0.91
7	Makassar	2010	32566	1.8	0.78
10	Bontoala	1184	26189	1.9	0.69
12	Ujung Tanah	639	8134	4.3	0.57
13	Tallo	2093	24867	4.7	0.46
4	Sangkarrang	38	9172	20	0.09

From Table 2, it can be seen that the first three highest relative risks of Covid-19 were Ujung Pandang (1.71), Rappocini (1.45), and Tamalanrea (1.39). Conversely, the first lowest relative risk of Covid-19 were Sangkarrang Island (0.09), Tallo (0.46), dan Ujung Tanah (0.57). The map of the relative risk of confirmed Covid-19 based on the preferred model can be seen in Figure 1 as follows.

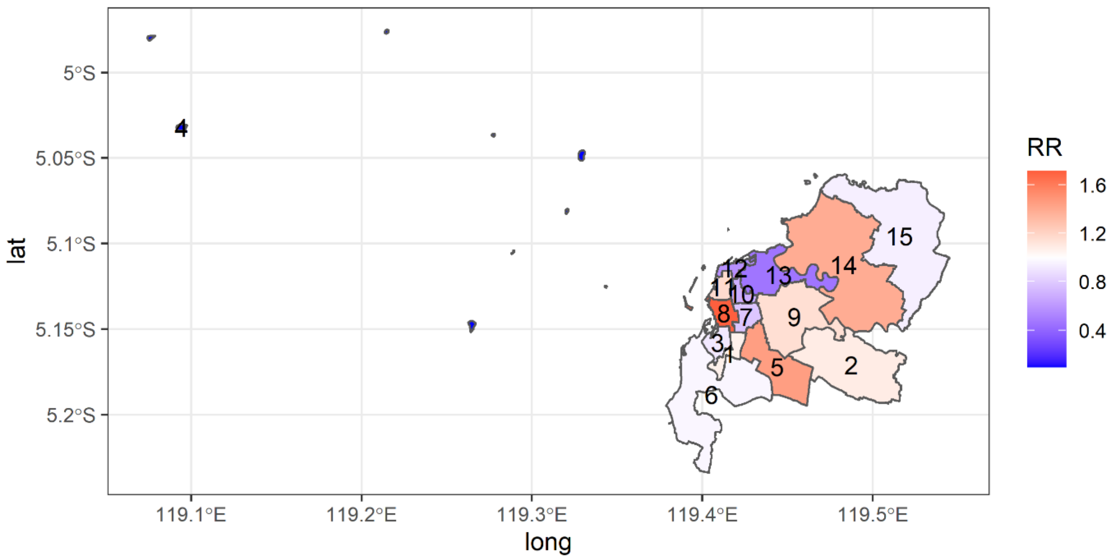

Figure 1. The relative risk map of confirmed covid-19 in Makassar based on the preferred model for each area

Based on Figure 1, Ujung Pandang as the capital city of Makassar city has the highest relative risk of Covid-19 cases. As Ujung Pandang is the administrative center of Makassar city [12], it can cause high human mobility within a district. This may cause Ujung Pandang to be the center of the virus Covid-19, but it needs further investigation. One study has found that Covid-19 incidence was strongly negatively associated with the distance from the virus epicenter [8]. On the contrary, Sangkarrang Island has the lowest relative risk of Covid-19. Based on Table 2, and Figure 1, Sangkarrang Island is located separately from other districts and the location is far from Ujung Pandang.

Overall, based on 95% CI, the smallest DIC and WAIC, and the closest MMI for residual to zero, the best model in modelling Covid-19 is spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) model with distance to the capital city incorporated. Our results found that distance to the capital city influences the relative risk of Covid-19. There was a negative correlation between the distance to the capital city and Covid-19 relative risk. It means that the closer a district is to the capital city, the greater the risk of being infected with Covid-19. This importance of distance to the capital city is somewhat similar to a study that there was a negative correlation between the Covid-19 incidence and the distance from the virus epicenter [8]. Furthermore, we also found that the association between population density and the relative risk of Covid-19 was not statistically significant. This result is in agreement with a study in Iranian Provinces [8].

4. Conclusion and future work

Our results indicate that it is recommended to try more than one model especially when the covariates were included in the model. Based on the results, it is concluded that the best model in modelling Covid-19 is spatial CAR Leroux with hyperprior Inverse-Gamma (0.5, 0.05) model with distance to the capital city incorporated. It is found that there was a negative correlation between the distance to the capital city and Covid-19 relative risk. It means that the closer a district is to the capital city, the greater the risk of being infected with Covid-19. We also found that the association between population density and the relative risk of Covid-19 was not statistically significant. Considering other covariates such as the human mobility index could be possible future work.

Acknowledgments
The authors wish to acknowledge the Universitas Negeri Makassar, Indonesia. This work was funded by DIPA Universitas Negeri Makassar Number: SP DIPA 023.17.2.677523/2021 November 2020, according to the Decree of the Rector of Universitas Negeri Makassar Number: 550/UN36/HK/2021 on May 05, 2021.

References
[1] Dinkes K M 2021 Info Penanggulangan Covid-19 Kota Makassar.
[2] CNN I 2020 Sulsel Jadi Wilayah Kasus Corona Tertinggi di Luar Pulau Jawa.
[3] Ntzoufras I 2009 *Bayesian modeling using WinBUGS* vol 698(New Jersey, USA: John Wiley & Sons)
[4] Gelman A 2013 *Bayesian Data Analysis, Third Edition*(Hoboken: CRC Press)
[5] Whittle R S and Diaz-Artiles A 2020 *BMC medicine* **18** 271-271
[6] DiMaggio C, Klein M, Berry C and Frangos S 2020 *Annals of epidemiology* **51** 7-13
[7] Konstantinoudis G, Padellini T, Bennett J, Davies B, Ezzati M and Blangiardo M 2021 *Environment international* **146** 106316
[8] Dadar M, Fakhri Y, Bjørklund G and Shahali Y 2020 *Archives of virology* **165** 2555
[9] Besag J, York J and Mollié A 1991 *Annals of The Institute of Statistical Mathematics* **43** 1-20
[10] Leroux B G, Lei X and Breslow N 2000 *Statistical Models in Epidemiology, the Environment, and Clinical Trials* **116** 179-191
[11] Aswi A, Cramb S, Duncan E and Mengersen K 2020 Int. J. Health Geogr. 19 39
[12] Statistik B P 2021 Makassar Municipality in Figures 2021. (Makassar: BPS)
[13] Lee D, Rushworth A and Napier G 2018 J. Stat. Softw. 84 1-39
[14] R Core Team 2019 R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing)
[15] Waller L A and Gotway C A 2004 *Applied spatial statistics for public health data* (Hoboken, N.J: John Wiley & Sons)
[16] Oyana T J and Margai F 2015 *Spatial analysis: statistics, visualization, and computational methods* (Boca Raton: CRC Press)
[17] Carrijo T B and Da Silva A R 2017 *Geogr. Anal.* **49** 451-467
[18] Aswi A, Cramb S, Duncan E and Mengersen K 2021 *Journal of physics. Conference series* **1899** 12098
[19] Moran P A P 1950 *Biometrika* **37** 17
[20] Spiegelhalter D J, Best N G, Carlin B P and Van Der Linde A 2002 *Journal of the Royal Statistical Society. Series B, Statistical methodology* **64** 583-639
[21] Watanabe S 2010 *J. Mach. Learn. Res.* **11** 3571-3594