Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 175
DOI: 10.11159/icsta25.175

Bivariate Normal Distribution for Modelling Spatio-Temporal Data

Rinda Nariswari[1,2], Kartika Fithriasari[1], Nur Iriawan[1]
[1]Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7003221004@student.its.ac.id; rinda.nariswari@binus.ac.id; kartika_f@its.ac.id; nur.iriawan@its.ac.id
[2]Statistics Department, School of Computer Science, Bina Nusantara University
Jakarta, Indonesia

Abstract – Spatio-temporal data is typically modelled using the shared spatial component and conditional autoregressive (CAR) models, with spatial random variables assumed to be normally distributed. To monitor the COVID-19 pandemic in Jakarta province, Indonesia, daily data on epidemiological indicators are collected for each community. Analysing the spatio-temporal patterns of these markers can shed light on the spread of COVID-19 in Jakarta province, Indonesia. While univariate spatio-temporal models have been widely researched, a joint model that considers both spatial and temporal effect is required to examine the relationship between different outcomes. We propose using a bivariate-normal spatio-temporal distribution in conditional autoregressive models. Our work aimed to create a bivariate spatio-temporal model to analyse the relationship among the number of positive confirmed COVID-19 cases and the number of deaths from COVID-19 in Jakarta province since 2020 to early 2024. The data consists of 42 municipalities and 199 weeks. However, we tried to fit the data with the normality assumption. We employed a bivariate normal conditional autoregressive model (BNCAR) to account for spatial pattern correlation. To account for temporal correlation, we used a bivariate random walk prior and a bivariate normal conditional autoregressive prior. The estimation follows a Bayesian framework. The Bayesian disease mapping approach using a bivariate normal conditional autoregressive model showed some areas for the number of confirmed COVID-19 cases and the number of deaths from COVID-19 in the Jakarta province area.

Keywords**:** COVID-19, Bayesian, bivariate, spatio-temporal modelling, disease mapping

1. Introduction

Spatio-temporal data, which have variations in both spatial and temporal dimensions, have gained prominence in many disciplines, including epidemiology, environmental science, urban planning, and geosciences. Spatio-temporal data present special challenges compared to data that are purely spatial or temporal in nature because of their intrinsic complexity, autocorrelation, and the requirement to model jointly evolving changes in space and time [1]. The combination of space and time frameworks enables a more precise explanation of real-world phenomena, for example, the diffusion of diseases, climatic variations, and traffic patterns, through the incorporation of local interdependencies and development in time.

The recent development in statistical modelling and computational techniques, specifically within the framework of the Bayesian hierarchical model, has significantly improved the ability to model and understand spatio-temporal data [2]. Such models offer flexible frameworks to incorporate several sources of uncertainty and established information, enabling strong inference and forecasting functions. Furthermore, recent advances in spatio-temporal modelling—exampled by dynamic Gaussian processes, conditional autoregressive (CAR) models, and Markov switching processes—have greatly enhanced the analytical tools at the disposal of researchers dealing with nonstationary and regime-changing phenomena [3], [4]. In spite of such advances, spatio-temporal data modelling is still a methodologically and computationally demanding task, particularly against the backdrop of complications such as non-Gaussian, high dimensionality, and missing or irregular observations. As a result, current research work is persistently seeking computationally efficient estimation methods as well as better model specifications that can uncover the complex relationships inherent in spatiotemporal data sets [5].

The COVID-19 pandemic has triggered an unprecedented demand for accurate and timely models to understand, monitor, and predict the spread of infectious diseases across space and time. Given the virus's highly dynamic and geographically heterogeneous nature, classical epidemiological models that consider either spatial or temporal aspects in isolation are often inadequate. Instead, spatio-temporal models have become essential in capturing the dual influence of

location and time on disease transmission, enabling a deeper understanding of how COVID-19 evolves within and across communities[6]; [7].

Spatio-temporal epidemiological modelling allows for fine-grained analysis of disease patterns, supports real-time surveillance, and informs targeted intervention strategies. Within this modelling framework, Bayesian hierarchical models have gained prominence due to their flexibility in incorporating structured spatial and temporal random effects, prior information, and various sources of uncertainty [2], [8]. Among the spatial modelling techniques, the Conditional Autoregressive (CAR) model has proven particularly effective for areal data, where regions are represented as discrete spatial units [9]; [10].

The CAR model introduces spatial dependence by assuming that the value of a parameter in a given area is conditionally dependent on the values in its neighbouring areas. This formulation is especially powerful in public health applications where administrative boundaries (e.g., districts, provinces) are the natural units of analysis and where spatial autocorrelation is expected due to population mobility, shared healthcare infrastructure, or socio-economic similarities [8]. In the context of COVID-19, CAR models help smooth the estimated risk across neighbouring regions, reducing noise from data sparsity or reporting variability and improving predictive performance.

Furthermore, when extended into a spatio-temporal setting, CAR models can be combined with autoregressive time components to model temporal persistence while leveraging spatial correlations [11]. Such spatio-temporal CAR models have been used to identify persistent high-risk clusters, detect outbreak hotspots, and evaluate the impact of intervention policies over time [12];[13]. They are also highly adaptable, allowing for integration with non-Gaussian likelihoods and covariate effects, which are common in infectious disease modelling.

This paper proposes a Bayesian spatio-temporal framework using a bivariate Normal CAR prior to model the spread of COVID-19 across Jakarta province, coupled with temporal dynamics to capture trends and fluctuations over the course of the pandemic. This manuscript aims to investigate and promote the creation of statistical models specific to spatiotemporal data with an emphasis on bivariate normal conditional autoregressive (BNCAR) model with Bayesian framework improving inference and prediction of regional COVID-19 risks, particularly under data uncertainty and spatial heterogeneity.

2. Data

We obtained data from March 2020 to January 2024, include the highest wave of COVID-19 infections and death in Jakarta province. This period also included two variants of viruses such as delta and omicron. The area divided into 42 municipalities (Fig.1) with range of cases from 0 to 30,800 cases per days. Daily confirmed and death case data were removed by municipality and aggregated by seven days (weekly) to prevent null number in observation and adjust for weekday and weekend impacts [14].

Fig. 1 Map of the 42 municipalities in Jakarta Province

3. Methodology

Let y_{ijk} is the number of confirmed cases COVID-19 (k= 1) and number of death COVID-19 (k= 2) in municipality i (i = 1,…,u = 42) at week j (j = 1,…,v = 199). To correct for weekday/weekend affects and excessive zero numbers in daily data, we modelled weekly data instead [14]. We assume that death COVID-19 cases and deaths have a Poisson distribution with a mean of μ_{ijk} [8]

$$y_{ijk} \sim \text{Poisson}\ (\mu_{ijk}) \tag{1}$$

3.1 Bivariate spatio-temporal models

In order to examine the relationship between spatial and temporal patterns, we need a joint model for bivariate data. Likewise univariate models, we assumed a Poisson distribution for each variable. The log relative risk is constructed based on spatial and temporal random effects.

$$y_{ij1}/\mu_{ij1} \sim \text{Poisson}\ (\mu_{ij1}) \tag{2}$$
$$y_{ij2}/\mu_{ij2} \sim \text{Poisson}\ (\mu_{ij2}) \tag{3}$$

The random effects are specified as follows. The u_{ik} and v_{ik} are the correlated and uncorrelated spatial effects, respectively, while g_{jk} is the temporal effects, and ψ_{ik} is the space-time interaction. Differ from univariate models, this study assumes a bivariate distribution for u_{i1} and u_{i2} and for g_{j1} and g_{j2} to capture the correlation between spatially random effects and temporal ceffects.

$$v_{ik} \sim \text{Normal}\ (0, \boldsymbol{\tau}_{v_k}{}^{-1}) \tag{4}$$
$$\psi_{ijk} \sim \text{Normal}\ (0, \boldsymbol{\psi}_k{}^{-1}) \tag{5}$$
$$g_{jk} \sim \text{BCAR}(1, \Lambda_{\boldsymbol{g}}) \tag{6}$$
$$u_{ik} \sim \text{BCAR}(1, \Lambda_{\boldsymbol{u}}) \tag{7}$$

The bivariate model was fitted using the Bayesian estimation, with MCMC methods employed to estimate posterior distributions. MCMC procedure was carried out utilising the NIMBLE software. We performed four scenarios of simulation with different number of samples was generated. The first simulation, we used 10,000 iterations with first 100 iterations was removed as burn-in. The second scenario, we used 100,000 iterations with first 1000 iterations of each chain was removed as burn-in. The third scenario, we used 1,100,000 iterations with first 100,000 iterations of each chain was removed as burn-in. For the last scenario, we used 3,300,000 iterations each. The first 300,000 iterations of each chain were removed as burn-in, and a thinning factor of 1000 was applied. For all standard deviations (σ_{uk}, σ_{vk}, σ_{gk}, $\sigma_{\psi k}$), defined in terms of the precision as $\sigma = \tau^{-1/2}$, a uniform distribution was over the range (0,10)[8], [15]. [11] provides code for NIMBLE that fits the MVCAR concept. Inverse-Wishart priors were implemented to $\boldsymbol{\Lambda_g}$ and $\boldsymbol{\Lambda_u}$ with 2 × 2 identity matrix as the scale matrix.

4. Result

4.1 Exploratory Data Analysis

We used graphical displays to explore the data prior to fitting models**.** To analyse the temporal trajectory of the two number of cases, we plotted them separately in Fig. 2 and Fig. 3 for each municipality throughout 199 weeks in our study. For this analysis, we got the insight that COVID-19 in Jakarta has two waves that impact increasing the number of confirmed cases, there happened in week 65 (30,800 cases) and 99 (23,480 cases). Cengkareng had the highest number of death (564), followed by Duren Sawit (521) and Grogol Petamburan (475). Fig.3 is a map of number of death across all weeks for each municipality in Jakarta which have been analysed by the author.

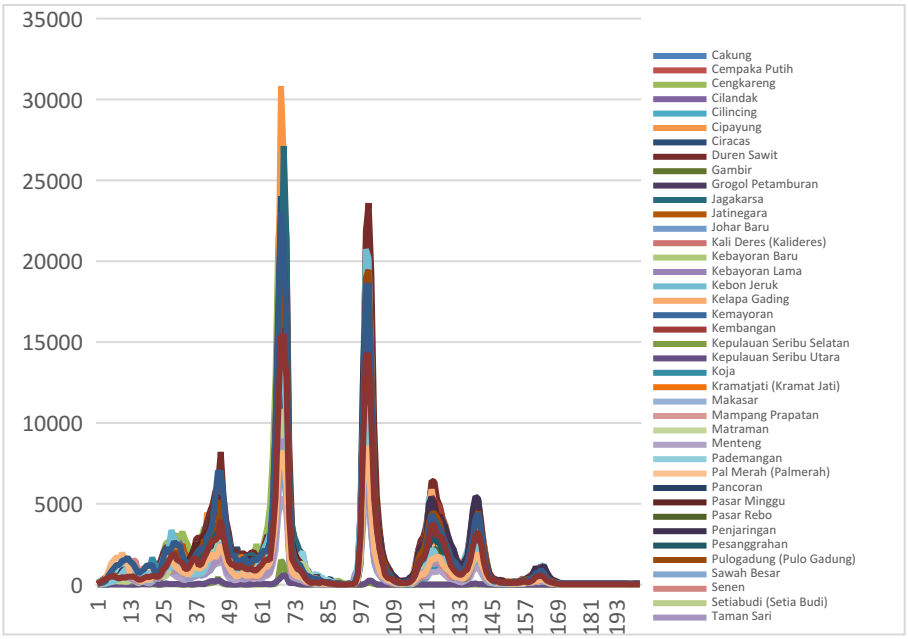


Fig. 2 Temporal Trend of Confirmed Cases for Each Municipalities in Jakarta

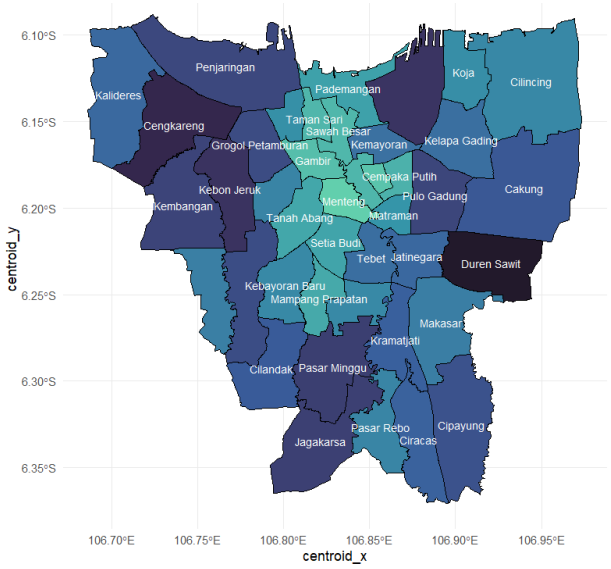


Fig. 3 Number of Death across all weeks for each municipality

4.2 Posterior Diagnostic Interpretation in a Spatio-Temporal COVID-19 Model

The findings in the MCMC diagnostic plots offer vital information regarding convergence, stability, and posterior parameter distributions from the Bayesian spatio-temporal method. Such diagnostics are essential in order to facilitate credible inference, particularly in spatial epidemiology, where uncertainty and heterogeneity in modelling must be accounted for in order to conduct effective disease surveillance and policymaking.

4.3 Convergence and Stability of MCMC Chains

Trace plots for all parameters show good convergence and mixing behaviour. Parameters fluctuate without discernible trends or drifts from one iteration to the next, which is a sign that the Markov chains have converged to their stationary distributions. This ensures that posterior samples are representative and summary statistics such as posterior means, medians, and credible intervals can be interpreted with confidence. Specifically, the lack of multimodality or poor mixing in trace plots suggests that the chains are not stuck in local modes, a desirable property when investigating complex dependencies in COVID-19 data over space and time.

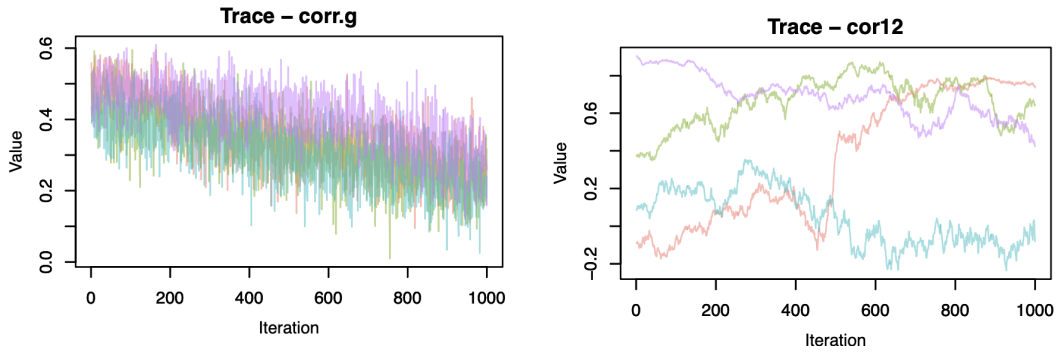


Fig.4 Correlation Terms using Trace Plot

Fig.4 explains these parameters, corr.g and cor12 likely capture spatial and temporal correlations among neighbouring regions or time points. Their posterior densities, centred around moderate positive values (e.g., 0.3 to 1.0), suggest strong dependency structures—consistent with how COVID-19 tends to cluster and propagate locally before spreading outward. Positive correlations imply that infection risks are not isolated events but are spatially or temporally linked, which justifies the use of a structured model such as the conditional autoregressive (CAR) prior.

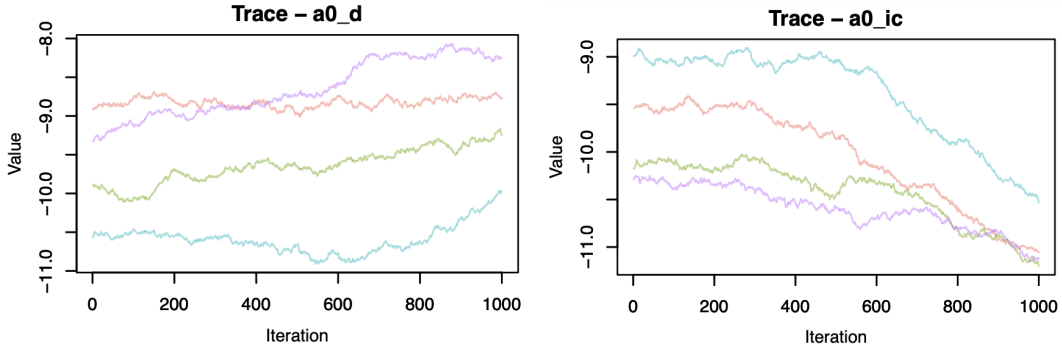


Fig. 5 Intercept Terms of Baseline Log-Risk Level

Fig.5 explains these negative-valued parameters likely represent baseline log-risk levels under different components (direct vs. indirect or structured vs. unstructured effects). Their stable posterior densities suggest a consistently low baseline infection rate**,** which can be interpreted as the background risk in the absence of strong spatial or temporal modifiers. In a COVID-19 context, this could reflect regions with lower mobility or more effective baseline interventions.

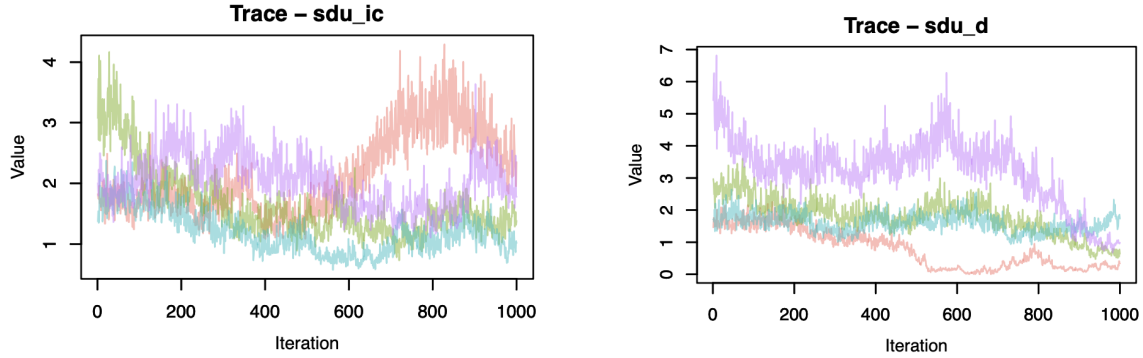


Fig. 6 Trace Plot of Structured Spatial Variability

Fig. 6 explains the standard deviations of the spatial random effects reveal the degree of spatial heterogeneity across the study area. For instance, sdu_d having a higher posterior spread indicates stronger spatial variation, possibly corresponding to clusters of high transmission such as urban centres, transportation hubs, or socioeconomically vulnerable areas. These parameters reinforce the importance of accounting for =

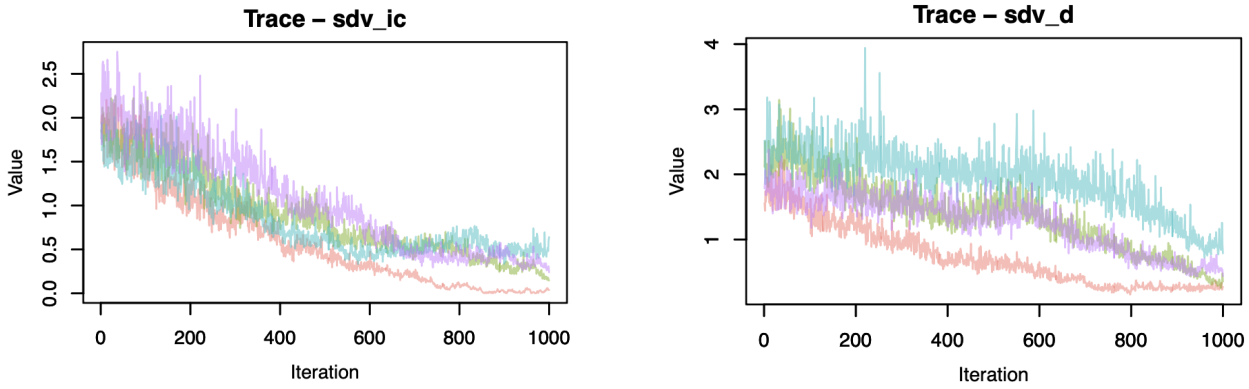


Fig. 7 Trace Plot of standard deviations account for residual spatial variability

Fig. 7 explains that these standard deviations account for residual spatial variability not captured by the structured CAR component. High values indicate non-systematic local deviations, potentially due to reporting inconsistencies, superspreader events, or other idiosyncratic factors. These components ensure that the model is flexible enough to capture both structured and random variation, which is essential in a pandemic where not all spread mechanisms are spatially or temporally predictable.

The combination of well-behaved posterior distributions and clearly interpretable parameters supports the robustness of the fitted spatio-temporal model. The application of CAR priors provides substantial benefits by smoothing risk estimates across neighbouring regions, thus improving the stability of estimates in areas with sparse or noisy data—common in early or underreported phases of the pandemic. From a policy standpoint, the spatial standard deviations help identify regions with elevated variability, potentially flagging them for targeted surveillance or intervention. Meanwhile, temporal parameters aid in projecting future trends or evaluating the lagged impact of public health measures such as lockdowns, school closures, or vaccination campaigns.

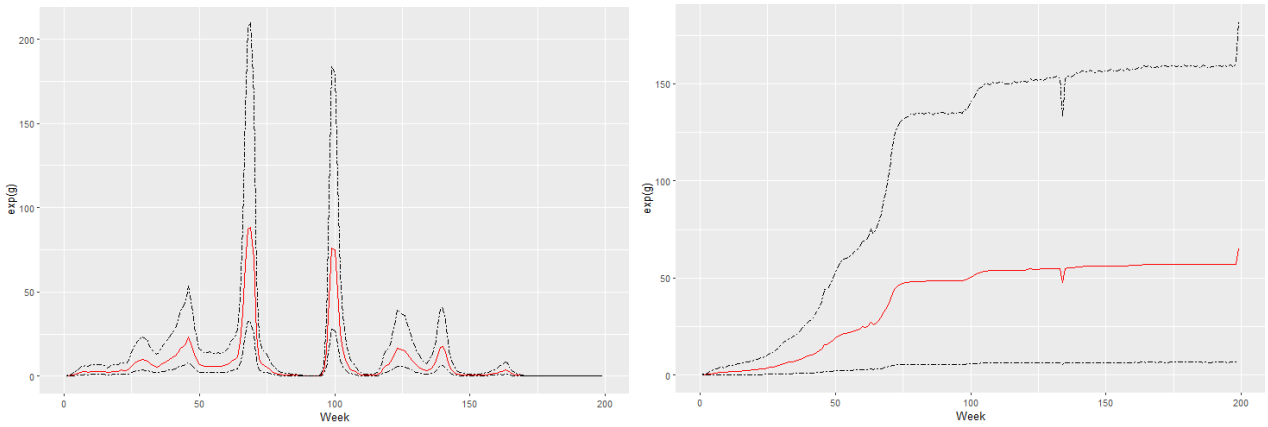
Overall, the diagnostics confirm that the Bayesian spatio-temporal model performs reliably and provides meaningful insights into the spatial and temporal patterns of COVID-19 transmission. The model's structure, incorporating both CAR priors and random effects, is well-suited to capture the complexity of pandemic spread. These results can be used to inform dynamic risk maps, resource allocation, and outbreak response strategies tailored to both spatial structure and temporal dynamics.

4.4 Result of Bivariate Models

Table 1 summarises the results from bivariate models. This table includes WAIC, pWAIC, temporal, and spatial correlations. The association between temporal random effects g_{j1} and g_{j2} was not significant for any of iteration's number, suggesting that events in time may affect both outcomes differently. The spatially organised random effects (u_{i1} and u_{i2}) showed a positive and substantial connection. Fig. 8 displays the estimated temporal trend of confirmed cases and death cases for bivariate model with the fourth scenario, which had the highest temporal correlation. We also displayed scenario 4 results as they had smallest WAIC and pWAIC. Fig. 9 shows the number of death cases, the geographical the trend indicates elevated spatial threats in the middle of Jakarta. The greatest spatial risks associated with death cases are observed in Cipayung

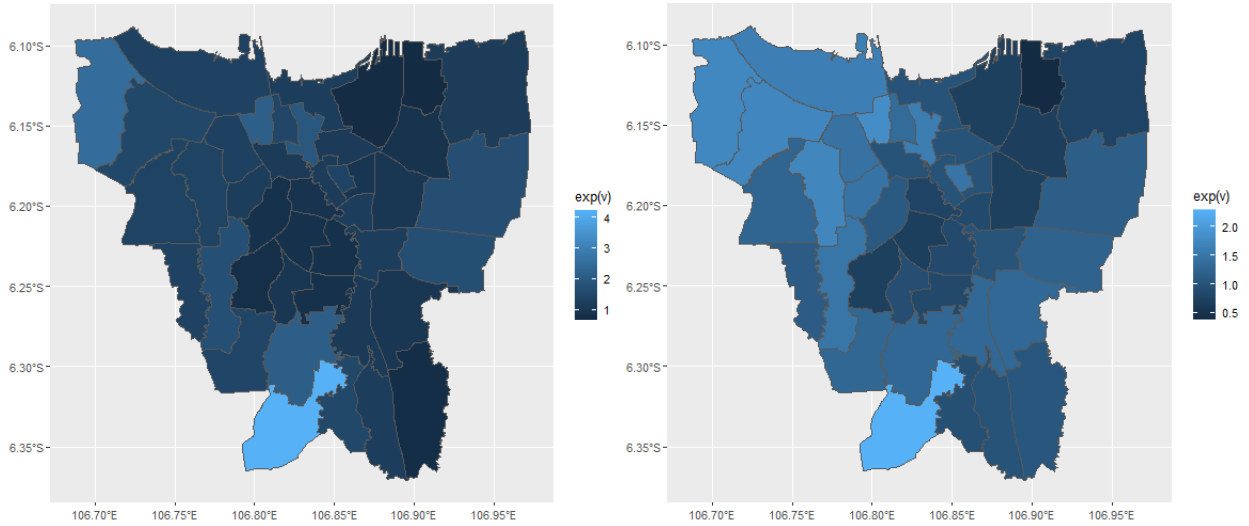
Table 1. Summary of the Result Bivariate Normal Model

Scenario	WAIC	pWAIC	Temporal Correlation	Spatial Correlation
1	185,015.2	8,477.22	0.3713 [-0.2997;0.5613]	0.3561 [0.1601;0.6211]
2	171,825.7	11,010.69	-0.0866 [-0.5545;0.3837]	0.3266 [0.1205;0.6424]
3	171,246.9	9,984.2	0.2072 [-0.1359;0.6012]	0.3650 [0.0685;0.6522]
4	171,212.7	9,909.89	0.1769 [-0.3535;0.5488]	0.3829 [0.0677;0.6304]



(a) Temporal Trend of Confirmed Cases (b) Temporal Trend of Death Cases

Fig. 8 Posterior Means of Temporal Risk



(a) Spatial Risk of Confirmed Cases (b) Spatial Risk of Death Cases

Fig. 9 Map of Spatial Risk COVID 19 in Jakarta

5. Conclusion

In the current study, confirmed case numbers and fatalities due to COVID-19 were studied for 42 Jakarta area using Spatio-temporal models were contemplated. In terms of the death statistics, the geographical of the pattern revealed high spatial risks in the northern, centre, and south parts of Jakarta. As for the number of death cases, the geographical the trend also indicates elevated spatial threats in the middle of Jakarta. The greatest spatial risks associated with death cases are observed in Cipayung. The temporal trend since the death cases have two peaks: one in week 65 and week 99. The climax of the worldwide temporal trend for the highest number of deaths is in week 65.

This paper aims to propose a framework for modelling joint spatio-temporal outcomes and analysing the relationship between confirmed and deaths cases. This manuscript study to examine the spatial and temporal relationship between these two variables. To do this, we used bivariate Normal models and the first use of an MCAR prior for temporal random effects. The studies show a substantial association between the number of confirmed and deaths cases, with higher risks in the centre of Jakarta and lower risks in the west. This suggests that shared spatial risk factors, such as demographic or socioeconomic activities, may impact both the number of confirmed and deaths cases.

Further research could explore potential extensions to the given models. Consider a multivariate proper CAR model instead of extending the ICAR prior for spatial random effects. A random walk prior of order 2 can be used instead of 1 for

temporal random effects. A multivariate Gaussian prior for random effects can be used to account for correlation among spatially unstructured effects. For space-time interactions, our models always used Gaussian priors that were independent. [11] presented a variety of interactions that could be investigated.

Data availability

All data used in this study is openly available in https://corona.jakarta.go.id.

Author Contribution

This study was compiled, designed, and analyzed by Rinda Nariswari and Kartika Fithriasari. Dataset and literature collection were prepared by Rinda Nariswari. The manuscript was reviewed by Kartika Fithriasari and Nur Iriawan. All authors have read and approved the manuscript.

References

[1] O. F. Christensen, "Statistics for Spatio-Temporal Data by CRESSIE, N. and WIKLE, C. K.," *Biometrics*, vol. 68, no. 4, hlm. 1328–1329, 2012, doi: https://doi.org/10.1111/j.1541-0420.2012.01835.x.

[2] S. Banerjee dan B. P. Carlin, *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. Boca Rotan, FL: CRC Press, 2014.

[3] F. Lindgren, H. Rue, dan J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *J R Stat Soc Series B Stat Methodol*, vol. 73, no. 4, hlm. 423–498, 2011, doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x.

[4] A. E. and S. A. M. and B. S. and S. C. F. Gelfand, "Spatio-temporal modeling for estate data," *J Econom*, vol. 134, no. 2, hlm. 119–164, 2005.

[5] M. J. Heaton *dkk.*, "A Case Study Competition Among Methods for Analyzing Large Spatial Data," *J Agric Biol Environ Stat*, vol. 24, no. 3, hlm. 398–425, Sep 2019, doi: 10.1007/s13253-018-00348-w.

[6] M. R. Desjardins, A. Hohl, dan E. M. Delmelle, "Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters," *Applied Geography*, vol. 118, hlm. 102202, 2020, doi: https://doi.org/10.1016/j.apgeog.2020.102202.

[7] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, dan N. A. Christakis, "Population flow drives spatio-temporal distribution of COVID-19 in China," *Nature*, vol. 582, no. 7812, hlm. 389–394, Jun 2020, doi: 10.1038/s41586-020-2284-y.

[8] A. B. Lawson, *Bayesian Disease Mapping*, 3rd ed. Boca Rotan, FL: Chapman and Hall/CRC, 2018. doi: 10.1201/9781351271769.

[9] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," 1974.

[10] N. Best, S. Richardson, dan A. Thomson, "A comparison of Bayesian spatial models for disease mapping," Februari 2005. doi: 10.1191/0962280205sm388oa.

[11] L. Knorr-Held, "Knorr-Held: Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk Bayesian Modelling of Inseparable Space Time Variation in Disease Risk," 1999.

[12] D. Lee, "A comparison of conditional autoregressive models used in Bayesian disease mapping", *Spatiotemporal Epidemiology*, vol. 2, no. 2, page 79–89, 2011, doi: https://doi.org/10.1016/j.sste.2011.03.001.

[13] P. Moraga, "Geospatial Health Data; Modeling and Visualization with R-INLA and Shiny; Edition 1." doi: 10.4324/9780429341823.

[14] S. K. Sahu dan D. Böhning, "Bayesian spatio-temporal joint disease mapping of Covid-19 cases and deaths in local authorities of England," *Spat Stat*, vol. 49, Jun 2022, doi: 10.1016/j.spasta.2021.100519.

[15] A. Gelman, D. A. Van Dyk, Z. Huang, dan W. J. Boscardin, "Transformed and parameter-expanded Gibbs samplers for multilevel linear and generalized linear models *," 2007. Tersedia pada: http://www.stat.columbia.edu/