Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 27 (1) : 27-32 (March - 2025)
https://doi.org/10.54386/jam.v27i1.2761
https://journal.agrimetassociation.org/index.php/jam

Research Paper

Spatio-temporal Bayes model for estimating the number of hotspots as an indicator of forest and land fires in Kalimantan Island, Indonesia

F. ROHIMAHASTUTI, A. DJURAIDAH* and H. WIJAYANTO

Department of Statistics and Data Science, IPB University, Bogor, West Java, Indonesia
**Corresponding Author: anikdjuraidah@apps.ipb.ac.id*

ABSTRACT

Forest and land fires often occur on the island of Kalimantan and have a widespread impact on neighboring countries. One indicator of forest and land fires is hotspot. Climate factors play an important role in determining hotspot patterns and trends in a location, which often fluctuate and are difficult to predict. This research aims to predict the number of hotspot spatially and temporally in the next month on Kalimantan Island and analyze the influence of local climate on hotspot events. The Bayesian Conditional Autoregressive method with Integrated Nested Laplace Approximation and optimal weight selection using Getis-Ord G are used to increase prediction accuracy. The distribution of hotspot is assumed to follow the Negative Binomial distribution. The research results show that the best model uses an additive approach and interaction with explanatory variables with a Deviance Information Criterion value of 97,799.8. Predictions from this model have a Root Mean Square Prediction Error of 7.08 and an Average Absolute Prediction Error of 0.63. However, the model still has limitations in predicting extreme events. Climatic factors such as low rainfall, long days without rain, high air temperatures, and low humidity contribute significantly to the increase in the number of hotspot in Kalimantan.

***Keyword*:** Bayesian Spatio-Temporal, Conditional Autoregressive, Hotspot, Forest Fire, Negative Binomial.

Large-scale forest and land fires events recorded in several years in recent past highlights the substantial impact of forest fires on Kalimantan (Arisman, 2020). A hotspot, defined as an area with higher temperatures than its surroundings, serves as an early indicator for detecting forest and land fires. Each hotspot is represented as a point with specific coordinates, and the total number of points, not their area, is counted for analysis. Detection is based on satellite data analysis at specific pixels, using algorithms that operate effectively under cloud-free atmospheric conditions (Giglio *et al.,* 2003). The ASEAN agreement sets a temperature threshold of 321°K (48°C) to classify a hotspot. While hotspots are essential for assessing fire risk, not all detected hotspot indicate active fires; field verification or supplementary data are often necessary to confirm actual fire occurrences. The accumulation of hotspot in a region correlates with an increased risk of fire (Putra *et al.,* 2018). Advances in remote sensing technology enable near real-time monitoring, making hotspot valuable indicators for rapid surveillance and early fire management over large areas.

Researchers have studied forest fires using various prediction methods such as polynomials and logistic functions (Nurdiati *et al.,* 2021), the copula (Najib *et al.,* 2022), and logistic regression (Mehta *et al.,* 2019). However, most of these studies focus solely on either spatial or temporal aspects, failing to fully capture the spatio-temporal characteristics of forest fires. Spatio-temporal, a technique for analyzing data with spatial and temporal dimensions, has been widely adopted by researchers for various applications (Ratnasari and Dewi, 2019). Rachmawati *et al.,* (2019) used an additive spatio-temporal Bayes approach with INLA to predict rainfall, while Hayati *et al.,* (2022) modeled the Tweedie compound Poisson Gamma distribution, and Djuraidah *et al.,* (2021) applied a generalized Pareto approach for extreme values. One of the techniques used in spatio-temporal modeling is Conditional Autoregressive (CAR) which was applied by Soroori *et al.,* (2019) for accident prediction and Djuraidah *et al.,* (2022) for analyzing Tuberculosis spread. The CAR method, known for its efficiency, continues to be developed, especially for applications in climatology, such as modeling forest and land fires, where geographic factors play a significant role.

***Article info - DOI*:** https://doi.org/10.54386/jam.v27i1.2761
Received: 30 September 2024; Accepted: 11 December 2024 ; Published online : 1 March 2025

This study addresses this gap by developing a spatio-temporal prediction model for hotspot in Kalimantan using the CAR method and a Bayesian approach with Integrated Nested Laplace Approximation (INLA), while examining the influence of climate factors. The model aims to provide more accurate predictions and support improved early warning systems for forest fire control in Kalimantan.

MATERIALS AND METHODS

Data collection

This study analyzes forest and land fires on Kalimantan Island of Indonesia using hotspot data and local climate parameters, including rainfall, number of days without rain, air temperature, humidity, and wind speed. The data spans from January 2006 to December 2020, with a consistent spatial resolution of 0.25° × 0.25° and a monthly time scale. Hotspot data were obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysics, which monitors the number of hotspots identified each month across the Asia-Pacific region. Precipitation, air temperature, and wind speed data were derived from the Reanalysis 5th Generation (ERA5) monthly average data at a single level and rainless days was taken from hourly data. Humidity was calculated based on the ERA5 monthly average data at the pressure level. All these data were obtained from the European Centre for Medium Range Weather Forecasts. The variables and units used include precipitation in meters, 2m temperature in °K, 10m wind speed in meters/second, precipitation in meters, and relative humidity in %.

Bayesian CAR with INLA approach

The Bayesian method is a statistical approach based on Bayes' theorem, differing from classical methods despite both using the likelihood function. In classical estimation, the objective is to maximize the likelihood function in relation to the parameters. On the other hand, the Bayesian approach considers all unknown parameters as random variables governed by prior distributions (Ntzoufras, 2009). By applying Bayes' theorem, the Bayesian method produces the joint posterior distribution of the parameters (King *et al.,* 2010) and derives estimates from the posterior marginal distributions, which are obtained by integrating the joint posterior distributions. Although this integration can be complex, the Bayesian method effectively addresses these challenges. The Bayesian spatio-temporal model in this study refers to the research (Knorr-Held, 2000) which is formulated as follows:

1. Linear model without explanatory variables

$$\eta_{it}=\beta_0+u_i+v_i+(\alpha+\delta_i)\times t$$

2. Additive model without explanatory variables

$$\eta_{it}=\beta_0+u_i+v_i+\gamma_t+\phi_t$$

3. Additive and interaction models without explanatory variables

$$\eta_{it}=\beta_0+u_i+v_i+\gamma_t+\phi_t+\delta_{it}$$

4. Linear model with explanatory variables (full model)

$$\eta_{it}=\beta_0+x_1\beta_1+x_2\beta_2+x_3\beta_3+x_4\beta_4+x_5\beta_5++u_i+v_i+(\alpha+\delta_i)\times t$$

5. Additive model with explanatory variables (full model)

$$\eta_{it}=\beta_0+x_1\beta_1+x_2\beta_2+x_3\beta_3+x_4\beta_4+x_5\beta_5+u_i+v_i+\gamma_t+\phi_t$$

6. Additive model and interactions with explanatory variables (full model)

$$\eta_{it}=\beta_0+x_1\beta_1+x_2\beta_2+x_3\beta_3+x_4\beta_4+x_5\beta_5+u_i+v_i+\gamma_t+\phi_t+\delta_{it}$$

y_{it} is the response variable at location and time t , while u_i is a spatially structured random effect, v_i is a spatially unstructured random effect, γ_t is a temporally structured random effect, ϕ_t is a temporally unstructured random effect, δ_i is the interaction effect of location and time in the linear trend model, and δ_{it} is the spatio-temporal interaction effect in additive model.

The posterior distributions of parameters in a hierarchical Bayesian model can be estimated using the Integrated Laplace Approximation (INLA) in R. This technique employs the Laplace approximation to approximate the marginal posterior distributions of the model's parameters. INLA is an analytical Bayesian inference method, applicable to complex and hierarchical additive models, including latent Gaussian variables (Rue *et al.,* 2009).

Model comparison and goodness test

The deviance information criterion (DIC), introduced by Spiegelhalter *et al.,* (2002), is a commonly used metric for evaluating model fit in Bayesian frameworks, extending the Akaike information criterion (AIC) to Bayesian model comparisons. DIC is composed of two parts: one assesses the model fit, while the other accounts for model complexity. The DIC is then defined as:

$$DIC=\bar{D}+p_D$$

In addition to the fit model criteria, the model with the best prediction will also be seen. The tests used are root mean square error prediction (RMSEP) and average absolute prediction error (AAPE) to measure the difference between the predicted value and the actual value which is defined as follows:

$$RMSEP=\sqrt{\frac{\sum_{i=1}^{S}\sum_{t=1}^{T}(y_{it}-\hat{y}_{it})^2}{S\times T}} \qquad AAPE=\frac{\left|\sum_{i=1}^{S}\sum_{t=1}^{T}(y_{it}-\hat{y}_{it})\right|}{S\times T}$$

RESULT AND DISCUSSION

Hotspot distribution

The Cullen-Frey graph in Fig. 1 illustrates the distribution of hotspot, identifying the Normal, Negative Binomial, and Poisson distributions as appropriate. Fig. 1 showing that the Negative Binomial is a better match than the Poisson. Determining the hotspot distribution based on the calculation of the variance-to-mean ratio (VMR), which produces a value of 54.29. A VMR value greater than 1 indicates overdispersion, a situation where the variance is greater than the mean (Fortin and Dale, 2005). This suggests greater variability in the data than predicted by the Poisson distribution, typically indicating clustering patterns. In the context of forest and land fires, a high VMR suggests that fires tend to occur simultaneously in specific locations, making the Negative Binomial distribution the most suitable choice for modeling the observed overdispersion and clustering.

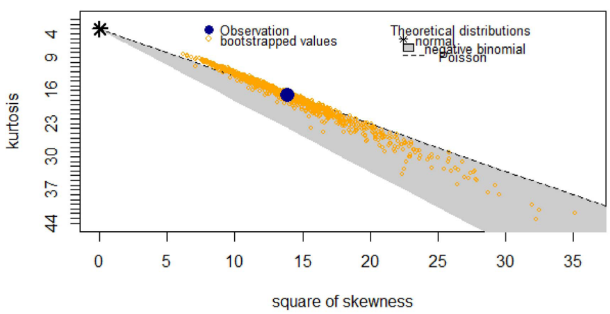

Fig. 1: Cullen and Frey graph

Table 1: Comparison of spatial weights based on standard global getis scores

Spatial Weight	Z(G)
k - Nearest Neighbor	28.5
Inverse Distance	38.9
Exponential Distance (α=1)	24.3
Exponential Distance (α=2)	24.1

Table 2: Model performance

Model	DIC	AAPE	RMSEP
Linear without explanatory variables	127538.2	1.6	6.8
Additive without explanatory variables	102764.3	1.2	8.3
Additive and interaction without explanatory variables	101971.3	0.9	5.6
Linear with explanatory variables	110627.5	6.7	203.1
Additive with explanatory variables	98914.5	1.2	8.1
Additives and interactions with explanatory variables	97799.8	0.8	4.9

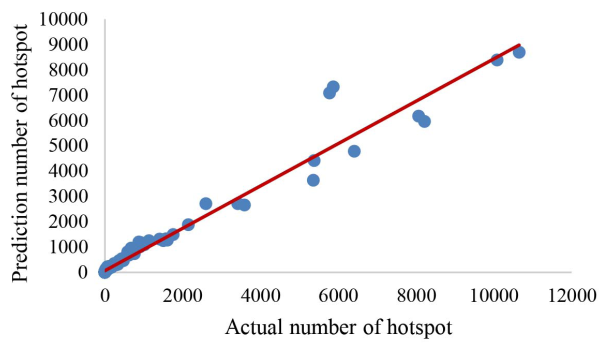

Fig. 2: Scatter plot between the actual number of hotspot and predictions from the best model in 2006

Optimal weight

Bayesian CAR modeling utilizes spatial dependency information to understand and predict agrometeorological phenomena, such as fire-prone areas influenced by climatic factors. In this study, selecting the optimal spatial weight is crucial for accurately modeling the distribution and risk of forest and land fires in Kalimantan. Table 1 shows that the inverse distance weight produces the highest standard Getis score, reflecting a clustering of high observation values. This finding highlights the importance of incorporating spatial autocorrelation to enhance the accuracy of predictions related to fire-prone zones, thereby supporting better mitigation strategies and resource management in agrometeorological contexts (Sukmawati *et al.,* 2021).

Spatio-temporal model

The spatio-temporal model, trained on data from 2006-2017. After prediction using the INLA method, six models were developed. As shown in Table 2, the best model based on the lowest DIC value is the spatio-temporal model with additive components and interactions with explanatory variables. The RMSEP for each model ranges from 5 to 8 hotspots, which is small compared to the actual average of 606 hotspots, indicating good predictive performance.

In Bayesian inference, posterior estimation relies on 95% credible intervals, which are generally determined by the 0.025 and 0.975 quantiles of the posterior distribution. Unlike classical confidence intervals, which assume symmetry, Bayesian models, as described by Perepolkin *et al.,* (2023), use quantile functions that do not rely on distribution shape assumptions. Since the hotspot data follows a skewed Negative Binomial distribution, classical confidence intervals are unsuitable, making quantile-based approaches crucial for accurate analysis. The estimation results for the best model are shown in Table 3.

The parameter β_0 is intercept. In general, the estimated mean value gives significant results which are marked by a positive/negative credible interval or do not pass through zero, except for the x_4 (wind speed). A negative coefficient for x_1 (rainfall) indicates more rainfall reduces hotspots, while a positive coefficient for x_2 (days without rain) suggests longer dry periods increase hotspots. Similarly, the positive coefficient for x_3 (air temperature) shows higher temperatures lead to more hotspots, and the negative coefficient for x_5 (humidity) suggests lower humidity increases hotspots.

According to Blangiardo and Cameletti (2015) the proportion of spatial and temporal diversity, it can be calculated using the formula:

$$s_u^2 = \frac{\sum_{i=1}^{n}(u_i - \bar{v})^2}{n-1} \qquad s_\gamma^2 = \frac{\sum_{i=1}^{n}(\gamma_i - \bar{\gamma})^2}{n-1}$$

$$frac_{spatial} = \frac{s_u^2}{s_u^2 + \sigma_v^2} \qquad frac_{temporal} = \frac{s_\gamma^2}{s_\gamma^2 + \sigma_\phi^2}$$

Table 3: Best estimated model parameter values

Parameter	Average	Standard deviation	Quantile 0.025	Quantile 0.5	Quantile 0.975	Mode
β_0 (intercept)	-34.93	1.67	-38.2	-34.93	-31.68	-34.93
β_1 (rainfall)	-69.28	8.06	-85.09	-69.28	-53.47	-69.28
β_2 (days without rain)	0.07	0.01	0.06	0.07	0.08	0.07
β_3 (air temperature)	1.29	0.04	1.19	1.29	1.38	1.29
β_4 (wind speed)	0.04	0.05	-0.06	0.04	0.15	0.04
β_5 (humidity)	-0.02	0.01	-0.04	-0.02	-0.01	-0.02
τ_u^2 (spatially structured)	0.39	0.04	0.35	0.39	0.49	0.36
τ_v^2 (spatially unstructured)	858.29	167.84	558.60	847.83	1215.50	836.57
τ_γ^2 (temporally structured)	0.64	0.08	0.51	0.63	0.81	0.62
τ_ϕ^2 (temporally unstructured)	6.33	1.24	4.30	6.19	9.16	5.89
τ_δ^2 (interaction spatial and temporal)	1.28	0.03	1.21	1.27	1.35	1.26

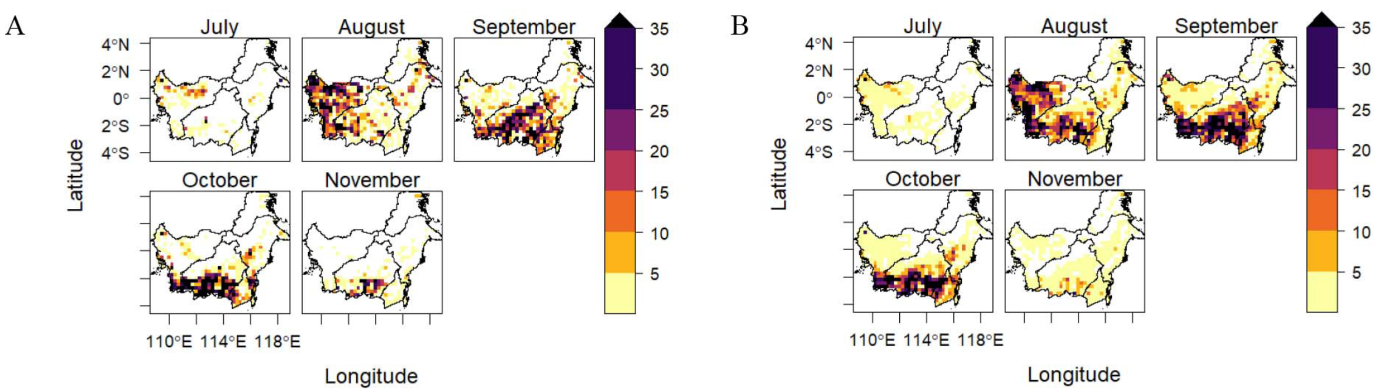

Fig. 3: Hotspot distribution map from the best model in 2006 for each month (A) actual (B) predicted

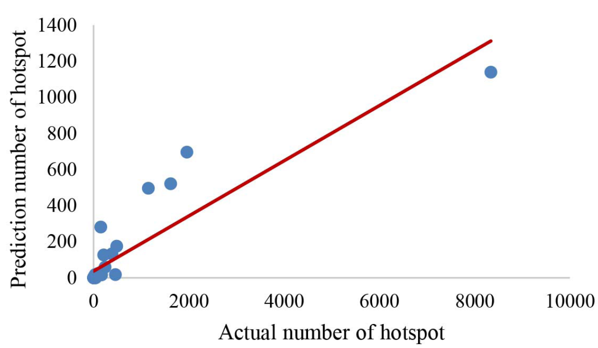

Fig. 4: Scatter plot between the actual number of hotspot and predictions from the 2018 model evaluation

The best model shows that 99% of the variation in the data is explained by spatial factors, while only 10% is due to temporal variation, indicating the data is stable over time but varies significantly by location. This stability is likely due to the use of monthly data, where the impact of forest and land fires doesn't change much from month to month. Shorter time intervals, like hourly or daily data, would likely show more dynamic fluctuations and significant impacts.

Fig. 2 shows most predictions tend to lie above this line, indicating that the model generally overestimates the number of hotspots. This suggests that the model is capable of capturing extreme spatial patterns, although there is a tendency for overprediction in some locations. Spatially, the model exhibits a similar pattern to the actual data, although the predictions tend to be higher. Fig. 3 displays the comparison between the distribution of actual hotspots and the predicted results on a monthly basis, successfully capturing the main spatial patterns. However, in some months, the predictions exceed the actual values, which may be attributed to overestimation in areas with a high fire risk.

Model evaluation

Model evaluation utilized testing data from 2018-2020 to assess performance. The results indicate that the prediction errors RMSEP of 7.08 and AAPE of 0.63 are similar to those from the training data, suggesting consistent model quality.

Fig. 4 shows that most of the data points are concentrated near the diagonal line for smaller hotspot counts (below 1,000), indicating that the model performs reasonably well for these cases. However, as the actual number of hotspots increases above 2,000, the predictions begin to deviate significantly below the diagonal line, showing a tendency of the model to underpredict the actual values. There is also an extreme outlier with an actual hotspot count exceeding 8,000, where the model's prediction remains

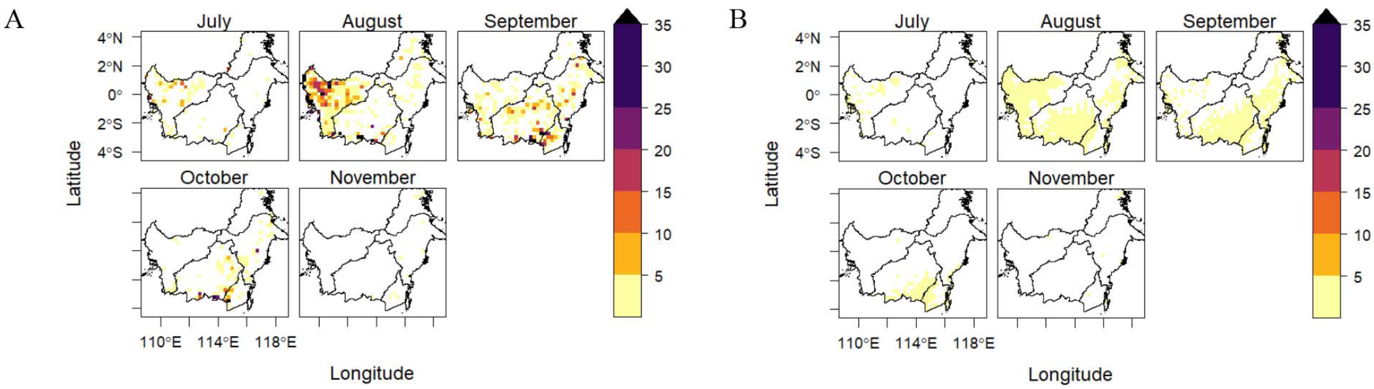

Fig. 5: Hotspot distribution map from the 2018 model evaluation results for each month (A) actual (B) predicted

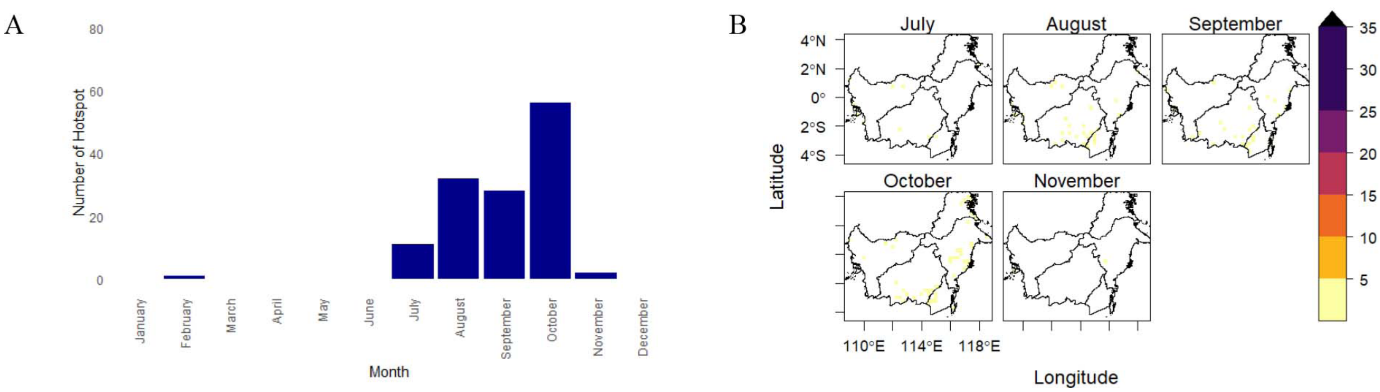

Fig. 6: Prediction of the number of hotspots in 2021 from (A) temporal (B) spatial aspects

much lower, indicating the model's difficulty in capturing patterns in extreme cases. This suggests that while the model is effective for predicting smaller-scale cases, it struggles to generalize for larger or more extreme hotspot counts. However, when viewed from the perspective of one location point in Fig. 5, the predicted number of hotspot is quite close to the actual value.

Estimation results

The results of the estimation using the additive and interaction models on the explanatory variables provide a potential picture of the number of hotspot in 2021, based on previously observed patterns. Predictions of the number of hotspot spatially and temporally can be seen in Fig. 6. Predictions from this model show a fairly high potential in October with 56 hotspot, followed by August with 32 hotspot, and September with 28 hotspot. Spatially, Central Kalimantan Province is the location with the highest potential for hotspot.

CONCLUSION

The optimal model for predicting hotspot occurrences is the additive and interaction model with explanatory variables, which achieved a DIC value of 97,799.8. The model demonstrates a low prediction error, as indicated by an RMSEP of 7.08 and an AAPE of 0.63. While it captures spatial and temporal hotspot patterns, its ability to predict extreme events remains limited based on model evaluation results. The model indicates a strong spatial variation (0.99) and a smaller temporal variation (0.10). Local climate factors, such as low rainfall, extended dry periods, high temperatures, and low humidity, are identified as key contributors to increased forest fire risk. However, further analysis is needed to assess the impact of wind speed on hotspot. This analysis contributes to understanding forest fire dynamics in Kalimantan and aids in developing effective mitigation strategies.

ACKNOWLEDGEMENT

The authors are grateful for the support of research funding from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia. We would also like to thank Indonesian Agency for Meteorology, Climatology and Geophysics for providing hotspot data.

Conflict of Interests: The authors declare that there is no conflict of interest related to this article.

Funding: This research was funded by the Directorate General of Higher Education, Research, and Technology of the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia in accordance with the contract for the implementation of the 2024 research program number 027/E5/PG.02.00.PL/2024 dated June 11, 2024.

Data availability: Dataset is provided in this URL: https://cds.climate.copernicus.eu/.

Authors contribution: **F. Rohimahastuti**: Conceptualization, methodology, investigation, writing-original draft preparation, visualization, editing; **A. Djuraidah**: Supervision, conceptualization,

methodology, investigation, writing-review, editing; **H. Wijayanto:** Supervision, conceptualization, methodology, investigation, writing-review, editing.

Disclaimer: The contents, opinions, and views expressed in the research article published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

***Publisher's Note*:** The periodical remains neutral about jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Arisman, A. (2020). Analisis Tren Kebakaran Hutan dan Lahan di Indonesia Periode Tahun 2015-2019. *J. Sains Teknol. Lingkungan,* 6 (1): 1-9. https://doi.org/10.29303/jstl.v6i1.131

Blangiardo, M., and Cameletti, M. (2015). Spatial and Spatio-temporal Bayesian Models with R-INLA. Chichester: John Wiley & Sons Ltd.

Djuraidah, A., Rachmawati, R.N., Wigena, A.H. and Mangku, I.W. (2021). Extreme data analysis using spatio-temporal Bayes regression with inla in statistical downscaling model. *Intern. J. Innov. Comp. Inform. Control,* 17 (1): 259-273. https://doi.org/10.24507/ijicic.17.01.259

Djuraidah, A., Mar'ah, Z. and Anisa, R. (2022). A Bayesian conditional autoregressive with Inla: A case study of tuberculosis in java, Indonesia. *Commun. Math. Biol. Neurosci.,* 2022: 1-15. https://doi.org/10.28919/cmbn/7709

Fortin, M.J. and Dale M.R.T. (2005). Spatial Analysis: A Guide for Ecologists. Cambridge University Press.

Giglio, L., Descloitres, J., Justice, C.O. and Kaufman, Y.J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sens. Environ.,* 87: 273-282. https://doi.org/10.1016/S0034-4257(03)00184-6

Hayati, M., Wigena, A.H., Djuraidah, A. and Kurnia, A. (2022). The spatio-temporal for the tweedie compound poisson gamma response in statistical downscaling. *J. Math. Comput. Sci.,* 12: 1-17. https://doi.org/10.28919/jmcs/6965

King, R., Morgan, B. J. T., Gimenez, O. and Brooks, S. P. (2010). Bayesian Analysis for Population Ecology. USA: Chapman & Hall/CRC.

Knorr-Held, L. (2000). Bayesian modeling of inseparable space-time variation in disease risk. *Stat. Med.,* 19 (17-18): 2555-67.

Mehta, D., Baweja, P.K. and Aggarwal, R.K. (2019). Developing fire danger model using logistic regression analysis for mid-hills of Himachal Pradesh. *J. Agrometeorol.,* 21 (4): 510-514. https://doi.org/10.54386/jam.v21i4.289

Najib, M.K., Nurdiati, S. and Sopaheluwakan, A. (2022). Copula-based joint distribution analysis of the ENSO effect on the drought indicators over Kalimantan fire-prone areas. *Model. Earth Syst. Environ.,* 8(2): 2817-2826. https://doi.org/10.1007/s40808-021-01267-5

Ntzoufras, I. (2009). Bayesian Modeling Using WinBUGS. New Jersey: Wiley.

Nurdiati, S., Sopaheluwakan, A., Julianto, M.T., Septiawan, P. and Rohimahastuti, F. (2021). Modeling and analysis of the impact of El Nino and IOD to land and forest fire using polynomial and generalized logistic function: case study in South Sumatra and Kalimantan, Indonesia. *Model. Earth Syst. Environ.*, 8(3): 3341-3356. https://doi.org/10.1007/s40808-021-01303-4

Perepolkin, D., Goodrich, B. and Sahlin, U. (2023). The tenets of quantile-based inference in Bayesian models. *Comp. Stat. Data Analysis,* 187. https://doi.org/10.1016/j.csda.2023.107795

Putra, I.D.G.A., Heryanto, E., Sopaheluwakan, A., Pradana, R.P. and Haryoko, U. (2018). Spatial and temporal distribution of hot spots in Indonesia from fashionable satellites using the grid-ding method. *Seminar Nasional Geomatika*. 3: 1123-1128. https://doi.org/10.24895/SNG.2018.3-0.1035

Rachmawati, R.N., Djuraidah, A., Wigena, A.H. and Mangku, I.W. (2019). Additive bayes spatio-temporal model with INLA for west Java rainfall prediction. *Procedia Comput. Sci.,* 157: 414–419. https://doi.org/10.1016/j.procs.2019.08.233

Ratnasari, N.R.P. and Dewi, V.R. (2019). Spatio-temporal model for predicting COVID19 cases in Indonesia. *Seminar Nasional Official Statistics*, 2020(1): 196-209. https://doi.org/10.34123/semnasoffstat.v2020i1.723

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B.,* 71: 319-392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Soroori, E., Moghaddam, A. M. and Salehi, M. (2019). Application of local conditional autoregressive models for development of zonal crash prediction models and identification of crash risk boundaries. *Transp. Sci.,* 15 (2): 1102-1123. https://doi.org/10.1080/23249935.2018.1564801

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J.R. Statist. Soc. B.,* 64 (4): 583. https://doi.org/10.1111/1467-9868.00353

Sukmawati S., Djuraidah, A. and Wigena, A.H. (2020). Spatial clustered regression analysis of 2017 getis score indonesian malaria prevalence data. *J. Phys. Conf. Ser.,* https://doi.org/10.1088/1742-6596/1863/1/012043