UCL DEPARTMENT OF GEOGRAPHY

YEAR 2024-25

EXAM CANDIDATE ID:	**LTKB8**
MODULE CODE:	**GEOG0125**
MODULE NAME:	**Advanced Topics in Social & Geographic Data Sciences**
COURSE PAPER TITLE:	**A Bayesian Spatial Approach to Risk Assessment: The Case of the 2020 Jakarta Flood**
WORD COUNT:	**1500 (excluding tables & figures contents and references)**
CODE REPOSITORY LINK:	**https://drive.google.com/drive/folders/1dB3X0Kcar-_Z3ENOuRSu8NBSJEZiuMYq?usp=sharing**

I acknowledge the use of Artificial Intelligence (AI) of ChatGPT (https://chatgpt.com/) to assist in improving my writing and troubleshooting code. I used ChatGPT to refine my writing for clarity and conciseness after drafting my initial ideas, ensuring that the core message and intent of my work remained intact. Additionally, I used ChatGPT to help troubleshoot coding issues. For instance, when I encountered difficulties removing units from a table in R, I used the prompt: “That is from ‘mutate(area = st_area(geometry)) %>% mutate(pop_density = Total.Population / area)’ and I want to remove the units and convert it to km².” To verify the output, I manually recalculated the values to ensure correct unit conversion. To maintain transparency, Moreover, all ChatGPT-assisted code sections were clearly credited within the submitted R script. My use of ChatGPT was primarily focused on improving the clarity of my writing and resolving specific coding issues.

Your essay, appropriately anonymised, may be used to help future students prepare for assessment. Double click this box to opt out of this ☐

1. Introduction

Floods are among the most destructive natural disasters, with long-lasting impacts on human life and infrastructure (Glago, 2021; Aldardasawi & Eren, 2021). Jakarta, identified as one of the top 20 global cities at highest flood risk by 2050, exemplifies this vulnerability (Wannewitz & Garschagen, 2021). In January 2020, the city experienced its worst flood in a decade, affecting over 100,000 residents and resulting in more than 60 fatalities (Nasution et al., 2022; Patel, 2020).

Flood prevention management in Jakarta is therefore a critical priority. With the growing availability of big data, data-driven analysis has significant potential in mitigating flood risks (Glago, 2021). Various studies have explored approaches to assess flood risk in Jakarta using big data techniques, including logistic regression (Hidayat Jati, Suroso & Santoso, 2019), Naïve Bayes classifiers (Saputra & Soetanto, 2024; Zamri, 2022), and machine learning methods (Haris et al., 2024; Priscillia, Schillaci & Lipani, 2022).

What distinguishes this study from previous research is its focus on estimating the number of people affected by the 2020 Jakarta flood using a multivariable Besag, York, and Mollié (BYM) model. Furthermore, the model will incorporate socioeconomic factors as independent variables. To the best of the author's knowledge, this is the first application of the BYM model for flood impact analysis in mainland Jakarta.

2. Data and Methods

Data for this study were primarily sourced from official Jakarta government reports, some were in PDF format and required manual extraction of relevant tables. The dependent variable is the number of flood-affected individuals per ward (Badan Penanggulangan Bencana Daerah, 2023). Independent variables include the Poverty-Prone Index and the Environmental and Health Vulnerability Index (Badan Pusat Data Statistik Provinsi DKI Jakarta, 2022). Population data at the ward level, used as an offset, were obtained from Dinas Kependudukan dan Pencatatan Sipil Provinsi DKI Jakarta (2021). A summary of all data sources is provided in Table 1.

Variables	**Details**	**Scholarly References**
Poverty-Prone Index	Indicator of poverty level	(McDermott, 2022; Tate et al., 2021)
Environmental and Health Vulnerability Index	Degree of waterway obstruction	(Zhou, 2014; Jha, Bloch and Lamond, 2012)
	Level of waste/garbage in the area	(Jha, Bloch and Lamond, 2012)
	Count of informal settlements along riverbanks	(Jha, Bloch and Lamond, 2012; Ferdous et al., 2020)

Table 1: Variable descriptions.

The BYM model that is used in this study consists of an Intrinsic Conditional Auto-Regressive (ICAR) term to account for spatial autocorrelation, and an unstructured random-effects term to capture non-spatial heterogeneity (Law, 2016; Duncan, White, & Mengersen, 2017). Due to its interpretability and computational efficiency, the BYM model is widely used in assessing excess risk across spatial units (Nayak et al., 2020; Quick, Song, & Tabb, 2021). The model is specified as follows:

$$
\begin{gathered}
Y_i \sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) = \alpha + \beta_1 X_1 + \beta_2 X_2 + C_i\sigma + \log(E_i) \\
\text{where } C_i = \theta_i + \phi_i = \left(\sqrt{1-\rho}\,\theta_i + \sqrt{\rho}\,\phi_i\right)
\end{gathered}
$$

Y_i is number of affected people by flood, E_i is the offset or expected numbers of affected people by flood in each ward. α is the baseline risk or the intercept, X_1 represents as the Poverty-Prone Index and X_2 represents as Environmental and Health Vulnerability Index and corresponding coefficients β_1, β_2 for each variables and σ is an overall error term multiplied to the combined random effects. Furthermore, θ_i is the ward-specific unstructured random effects, ϕ_i is the ward-specific spatial random effects, and ρ is the variance proportion of θ_i and ϕ_i. Lastly, $\exp(\alpha)$ is the global baseline risk ratio, $\exp(\beta)$ is the overall risk ratio for the coefficients, and $\exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + C_i\sigma)$ is the risk ratio for each wards (Musah, 2025).

To balance the contributions of the spatial and non-spatial components, gamma priors with parameters $\beta(0.5, 0.5)$ are applied to ρ, ensuring equal emphasis on θ_i and ϕ_i (Morris et al., 2019). To regularize the model, weakly informative priors of $\mathcal{N}(0,1)$ are used for α, β_1, β_2, and σ (Lemoine, 2019; Gelman, 2025; Li et al., 2022; Nayak et al., 2020; Musah, 2025). The model is estimated using Markov Chain Monte Carlo (MCMC) methods, with 4,000 iterations across 4 chains only, due to limited computational resources (Wang, 2022; Keefe, Ferreira & Franck, 2018).

3. Results

Model results indicated that the average number of people affected by the 2020 Jakarta flood was –3.76, with a relative risk of 0.03 (95% CrI: 0.01–0.08). This suggests that flood risk was approximately 97.67% lower in Jakarta that year. The result is statistically significant, with a near-zero probability of excess risk (See Table 2).

Parameter	**mean**	**exp(mean)**	**exp(2.5%)**	**exp(97.5%)**	**n_eff**	**Rhat**
α	-3.76	0.0282	0.0071	0.0791	2,746	1.001
β_1	-0.04	0.9631	0.9205	1.0028	2,139	1.000
β_2	-0.06	0.9387	0.8798	0.9959	2,518	1.000
σ	5.59	267.52	133.15	592.17	1,431	1.000

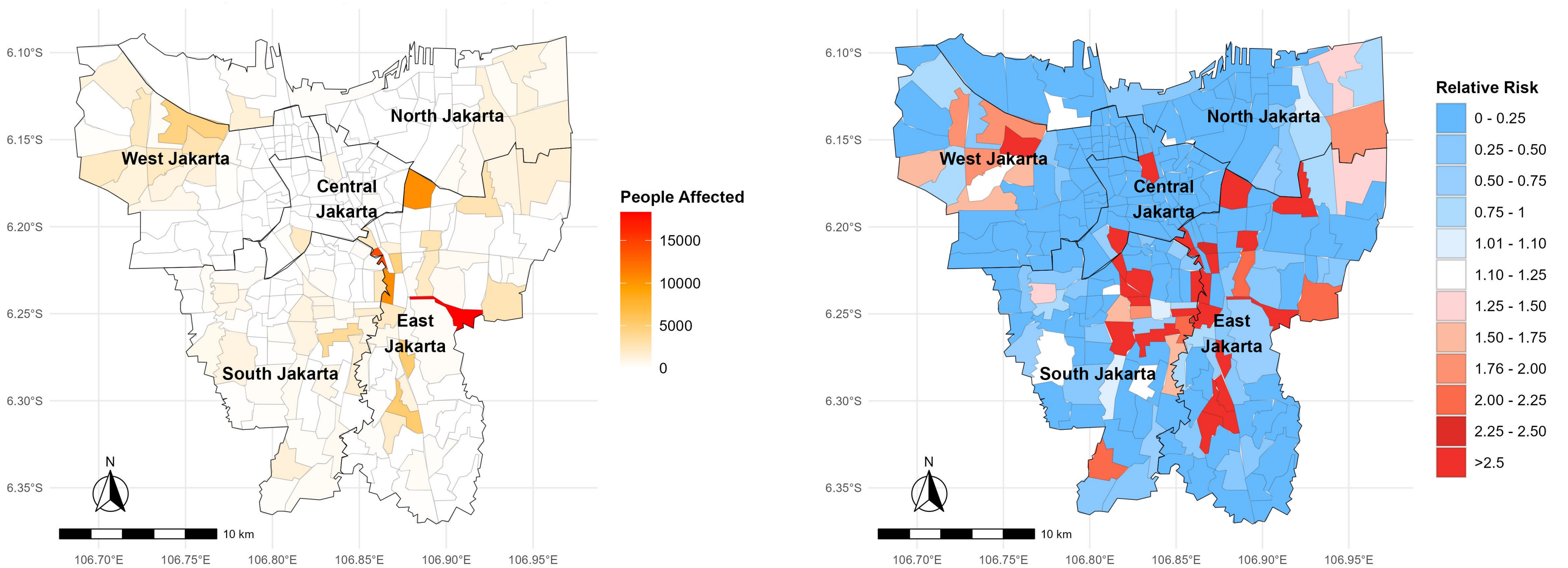
Table 2: Model summary statistics

Regarding the poverty index, areas with higher poverty levels showed a 0.038 average decrease (95% CrI: 0.92–1.00) in flood-affected populations. This corresponds to a relative risk of 0.96 (4% lower), with a 3% probability of excess risk. However, this result is not statistically significant, as the credible interval includes the null value (RR = 1).

For the environmental and health vulnerability index, the analysis shows an average decrease of 0.063 (95% CrI: 0.88–0.99), with a relative risk of 0.94 (6% lower) and a 3% probability of excess risk. This suggests that individuals residing in wards with lower environmental vulnerability may have a slightly higher likelihood of being affected by floods. Although statistically significant, the result is close to the threshold of non-significance.

At first glance, the results appear counterintuitive. Although 2020 was one of Jakarta's most severe flood years, the model indicates a 97% lower risk of flood impact on the population, particularly in areas with higher poverty and environmental vulnerability indices. These findings should be interpreted as exploratory. While they suggest that, at a broader scale, Jakarta exhibited lower flood-related risk and reduced exposure in socio-environmentally vulnerable areas, local variations reveal a more complex picture. This highlights the potential influence of other confounding factors that are absent in this research.

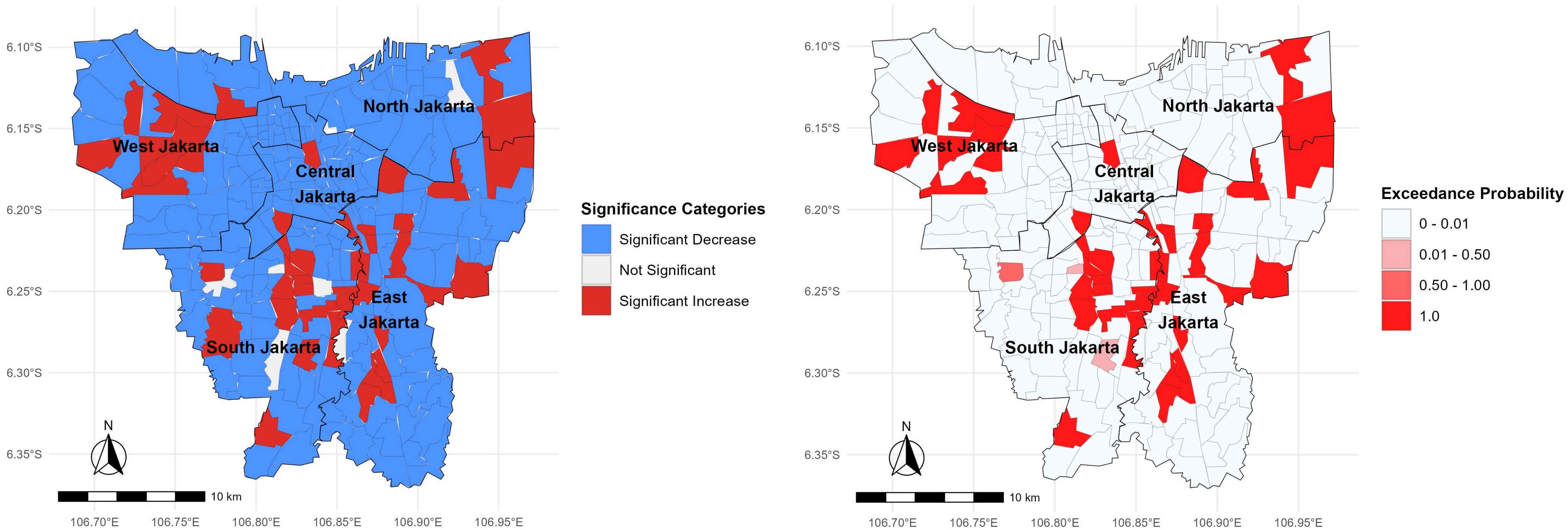
Figure 1a and figure 1b further highlights that the areas with the highest case and relative risk are primarily situated in East Jakarta. Furthermore, figure 2a reinforced that the majority of wards experienced a statistically significant decrease in flood risk. Moreover, a local threshold of 1.3 (or 30% higher risk than the baseline) was applied for exceedance probability. This threshold is based on research by Ban et al. (2023), who reported a relative risk of 1.3 for flood-related mortality in North Carolina, USA. When this benchmark is applied, both Figure 2b and Table 3 show that areas with high exceedance probabilities for relative risk more than 1.3 are notably concentrated in East and South Jakarta.



(a) Number of people affected by flood in Jakarta (2020)

(b) Relative risk of flood-affected population in Jakarta

Figure 1: Maps of number flood-impacted people and relative risk.



(a) Statistically Significance Map

(b) Exceedance Probability Map for Relative Risk 1.3

Figure 2: Significance and exceedance probability maps.

District	Number of Wards		
	Relative Risk > 1.3	Statistically Significant Increase	Statistically Significant Increase & Exceedance Probabilities >50% for 1.3 Relative Risk
East Jakarta	17	17	17
South Jakarta	14	16	14
West Jakarta	7	8	7
Central Jakarta	2	2	2
North Jakarta	2	3	2

Table 3: Number of high risk wards per Jakarta districts.

This study offers exploratory insights that highlights the need to reassess perceived flood risk. From the result of this study, wards considered low-risk based on low poverty levels or low environmental vulnerability indicators may, in fact, require urgent attention. While the independent variables used in this study are commonly associated with increased flood risk, the findings suggest they did not significantly explain flood risk in the case of the 2020 Jakarta flood. That year, extreme rainfall led to river overcapacity, and the absence of adequate bypass infrastructure along Jakarta's major rivers likely exacerbated flood impacts (Prabowo & Meiliana, 2020; Tempo, 2020; Gatra, 2020).

Although the environmental vulnerability index included the number of houses built along rivers, it may have overlooked critical factors such as river type and infrastructure quality, limiting its explanatory power. Consequently, even higher-income wards experienced substantial flood risk perhaps due to infrastructure constraints and widespread rainfall. In the specific context of the 2020 Jakarta flood, parts of East and South Jakarta would benefit from targeted interventions, including improved drainage systems, enhanced river and green infrastructure, and comprehensive flood prevention measures.

4. Conclusion

In this study, the author found that the global estimate of relative risk for flood-affected populations in Jakarta in 2020, as modeled using the BYM ICAR approach, was 0.03 (or 97.67% lower from the baseline). This is particularly notable given that 2020 recorded the highest number of people affected by floods. However, the study has several limitations.

First limitation is that the variables were sourced from official datasets published by Indonesian government agencies which are known to sometimes contain data inconsistencies and inaccuracies (Rachmat et al., 2024; Tjondronegoro et al., 2022). Second, the analysis is limited to the year 2020 and focuses only on global statistics. Future research could benefit from the use of Bayesian model updating or spatio-temporal models to assess changes in flood risk over time, as 2020 may represent an outlier. Additionally, spatially varying coefficient models (Meehan et al., 2024; Gelfand et al., 2003) could help capture local relationships between variables, particularly since this study identified wards with high relative risk statistics that are not being reflected in the global estimates. Third, several known confounding factors such as distance to rivers, rainfall intensity, elevation, and land use were not included in the model (Lin & Billa, 2021; Erhardt, Boudreault, & Carozza, 2022; Yusya, Septyandy, & Indra, 2020). Lastly, the use of aggregated ward-level data may introduce ecological bias, as not all locations within high-risk areas share the same level of vulnerability.

Despite these limitations, this research serves as an important exploratory analysis, offering insight into spatial patterns of flood risk in Jakarta and laying the groundwork for future research and policy development.

References

Aldardasawi, A.F.M. and Eren, B. (2021). *Floods and Their Impact on the Environment.* Academic Perspective Procedia, 4(2), pp.42–49. https://doi.org/10.33793/acperpro.04.02.24

Badan Penanggulangan Bencana Daerah (2023). *Data Kejadian Bencana Banjir di Provinsi DKI Jakarta Tahun 2020.* [online] Satu Data Jakarta.https://satudata.jakarta.go.id/open-data/detail?kategori=dataset&page_url=data-kejadian-bencana-banjir-di-provinsi-dki-jakarta-tahun-2020

Badan Pusat Data Statistik Provinsi DKI Jakarta (2022). *INDEKS POTENSI KERAWANAN SOSIAL PROVINSI DKI JAKARTA 2020.* [online] BPS Jakarta.https://jakarta.bps.go.id/id/publication/2022/06/09/e87d41b4f941968d011c2b60/indeks-potensi-kerawanan-sosial-provinsi-dki-jakarta-2020-.html

Ban, J., Sutton, C., Ma, Y., Lin, C. and Chen, K. (2023). *Association of flooding exposure with cause-specific mortality in North Carolina, United States.* Nature Water, 1(12), pp.1027–1034. https://doi.org/10.1038/s44221-023-00167-5

Dinas Kependudukan dan Pencatatan Sipil Provinsi DKI Jakarta (2021). *Buku Informasi Hasil Pelayanan Pendaftaran Penduduk dan Pencatatan Sipil di Provinsi DKI Jakarta.* [online] https://drive.google.com/file/d/1xoAJCuZeZ4FJaihsfIeBnX-tEvYfUC0h/view

Duncan, E.W., White, N.M. and Mengersen, K. (2017). *Spatial smoothing in Bayesian models: a comparison of weights matrix specifications and their impact on inference.* International Journal of Health Geographics, 16. https://doi.org/10.1186/s12942-017-0120-x

Erhardt, R.J., Boudreault, M. and Carozza, D.A. (2022). *Climate, Spatial Dependence, and Flood Risk: A U.S. Case Study.* Casualty Actuarial Society.

Ferdous, M.R., Di Baldassarre, G., Brandimarte, L. and Wesselink, A. (2020). *The interplay between structural flood protection, population density, and flood mortality along the Jamuna River, Bangladesh.* Regional Environmental Change, 20(1). https://doi.org/10.1007/s10113-020-01600-1

Gatra, S. (2020). *170 KK di Bidara Cina Terdampak Banjir, Genangan Sempat Setinggi 1,5 Meter.* [online] KOMPAS.com.https://megapolitan.kompas.com/read/2020/10/05/12213501/170-kk-di-bidara-cina-terdampak-banjir-genangan-sempat-setinggi-15-meter

Gelfand, A.E., Kim, H.-J., Sirmans, C.F. and Banerjee, S. (2003). *Spatial Modeling With Spatially Varying Coefficient Processes.* Journal of the American Statistical Association, 98(462), pp.387–396. https://doi.org/10.1198/016214503000170

Gelman, A. (2025). *Prior Choice Recommendations.* [online] Github. https://github.com/stan-dev/stan/wiki/prior-choice-recommendations

Glago, F.J. (2021). *Flood disaster hazards; causes, impacts and management: a state-of-the-art review.* Natural hazards–impacts, adjustments and resilience.

Haris, R., Haryo, W., Pujiharto, E.W., Yuza, A., Kusrini, K. and Kusnawi, K. (2024). *Prediksi Banjir Di Dki Jakarta Dengan Menggunakan Algoritma K-Means Dan Random Forest.* Jurnal Informatika dan Teknologi Komputer (J-ICOM), 5(1), pp.43–49.

Jha, A.K., Bloch, R. and Lamond, J. (2012). *Cities and flooding: a guide to integrated urban flood risk management for the 21st century.* The World Bank.

Keefe, M.J., Ferreira, M.A.R. and Franck, C.T. (2018). *On the formal specification of sum-zero constrained intrinsic conditional autoregressive models.* Spatial Statistics, 24, pp.54–65. https://doi.org/10.1016/j.spasta.2018.03.007

Law, J. (2016). *Exploring the Specifications of Spatial Adjacencies and Weights in Bayesian Spatial Modeling with Intrinsic Conditional Autoregressive Priors in a Small-area Study of Fall Injuries.* AIMS Public Health, 3(1). https://doi.org/10.3934/publichealth.2016.1.65

Lemoine, N.P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. Oikos, 128(7), pp.912–928. https://doi.org/10.1111/oik.05985

Li, L., Musah, A., Thomas, M.G. and Kostkova, P. (2022). An ecological study exploring the geospatial associations between socioeconomic deprivation and fire-related dwelling casualties in the England (2010–2019). Applied Geography, 144. https://doi.org/10.1016/j.apgeog.2022.102718

Lin, J.M. and Billa, L. (2021). Spatial prediction of flood-prone areas using geographically weighted regression. Environmental Advances, 6. https://doi.org/10.1016/j.envadv.2021.100118

McDermott, T.K.J. (2022). Global exposure to flood risk and poverty. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-30725-6

Meehan, T.D., Krainski, E.T., Lindgren, F. and Rue, H. (2024). Spatially Varying Coefficient Models with inlabru. [online] Github.io. https://inlabru-org.github.io/inlabru/articles/svc.html

Morris, M., Wheeler-Martin, K., Simpson, D., Mooney, S.J., Gelman, A. and DiMaggio, C. (2019). Bayesian hierarchical spatial models: Implementing the Besag York Mollié model in stan. Spatial and Spatio-temporal Epidemiology, 31. https://doi.org/10.1016/j.sste.2019.100301

Musah, A. (2025). BAYESIAN SPATIAL RISK MODELLING. UCL Department of Geography.

Nasution, B.I., Saputra, F.M., Kurniawan, R., Ridwan, A.N., Fudholi, A. and Sumargo, B. (2022). Urban vulnerability to floods investigation in jakarta, Indonesia: A hybrid optimized fuzzy spatial clustering and news media analysis approach. International Journal of Disaster Risk Reduction, 83. https://doi.org/10.1016/j.ijdrr.2022.103407

Nayak, M.A., Cowles, M.K., Villarini, G. and Wafa, B.U. (2020). Bayesian Hierarchical Models for the Frequency of Winter Heavy Precipitation Events Over the Central United States: The Role of Atmospheric Rivers. Water Resources Research, 56(11). https://doi.org/10.1029/2020wr028256

Patel, K. (2020). Torrential Rains Flood Indonesia. [online] NASA Earth Observatory. https://earthobservatory.nasa.gov/images/146113/torrential-rains-flood-indonesia

Prabowo, D. and Meiliana, D. (2020). *Ketua RW 03 Cipinang Melayu: Banjir Kali Ini Paling Besar dan Dua Kali.* [online] KOMPAS.com.https://nasional.kompas.com/read/2020/02/25/ 15425101/ketua-rw-03-cipinang-melayu-banjir-kali-ini-paling-besar-dan-dua-kali

Priscillia, S., Schillaci, C. and Lipani, A. (2022). Flood susceptibility assessment using artificial neural networks in Indonesia. Artificial Intelligence in Geosciences, 2, pp.215–222. https://doi.org/10.1016/j.aiig.2022.03.002

Quick, H., Song, G. and Tabb, L.P. (2021). Evaluating the informativeness of the Besag-York-Mollié CAR model. Spatial and Spatio-temporal Epidemiology, 37. https://doi.org/10.1016/j.sste.2021.100420

Rachmat, D.F., Mulyadi, D., Abdullah, S., Putrianti, S.D. and Nurliawati, R. (2024). Analysis of the Governance of One Data Indonesia (SDI) at the Communication and Information Office of Sukabumi City. In: In Fourth International Conference on Administrative Science (ICAS 2022). Atlantis Press, pp.42–60. https://doi.org/10.2991/978-2-38476-104-3_6

Saputra, S.A. and Soetanto, H. (2024). IMPLEMENTASI NAÏVE BAYES UNTUK KLASIFIKASI PREDIKSI SERTA ANALISIS DATA BANJIR DI WILAYAH JAKARTA PUSAT. In Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), 3(2), pp.296–304.

Tate, E., Rahman, M.A., Emrich, C.T. and Sampson, C.C. (2021). Flood exposure and social vulnerability in the United States. Natural Hazards, 106. https://doi.org/10.1007/s11069-020-04470-2

Tempo (2020). *Sodetan Kali Sunter Sudah Jadi, Cipinang Melayu Bebas Banjir.* [online] Tempo.https://www.tempo.co/arsip/ sodetan-kali-sunter-sudah-jadi-cipinang-melayu-bebas-banjir--576403

Tjondronegoro, D., Liew, A.W.-C., Verhelst, T., Green , D., Bernot , A., Hasan , R. and Rifai , B. (2022). The state of open data implementation in Indonesia. Regional Outlook Paper.

Wang, G. (2022). Laplace approximation for conditional autoregressive models for spatial data of diseases. MethodsX, 9. https://doi.org/10.1016/j.mex.2022.101872

Wannewitz, M. and Garschagen, M. (2021). Review article: Mapping the adaptation solution space – lessons from Jakarta. Natural Hazards and Earth System Sciences, 21(11). https://doi.org/10.5194/nhess-21-3285-2021

Yusya, R.R., Septyandy, M.R. and Indra, T.L. (2020). Flood Risk Mapping of Jakarta Using Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST) Methods. In: IOP Conference Series: Materials Science and Engineering.

Zamri, D. (2022). Perbandingan Metode Data Mining untuk Prediksi Banjir Dengan Algoritma Naïve Bayes dan KNN: Comparison of Data Mining Methods for Predition of Floods with Naïve Bayes and KNN Algorithm. SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat, pp.40–48.

Zhou, Q. (2014). A review of sustainable urban drainage systems considering the climate change and urbanization impacts. Water, 6(4), pp.976–992.