Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Sensitivity analysis of the PC hyperprior for range and standard deviation components in Bayesian Spatiotemporal high-resolution prediction: An application to PM2.5 prediction in Jakarta, Indonesia

Tafia Hasna Putri[a], I Gede Nyoman Mindra Jaya[b*], Toni Toharudin[c] and Farah Kristiani[d]

[a]*Department of Statistics, Padjadjaran University, Jl. Raya Bandung Sumedang km 21 Jatinangor, Sumedang 45363, Indonesia*
[b]*Department of Mathematics, Parahyangan University, Jl. Ciumbuleuit No. 94, Hegarmanah, Kec. Cidadap, Kota Bandung 40141, Indonesia*

C H R O N I C L E

Article history:
Received: November 10, 2023
Received in revised format: November 20, 2023
Accepted: December 18, 2023
Available online: December 18, 2023

Keywords:
Spatiotemporal Modeling
GMRF
Penalized Complexity (PC)
PM2.5 Concentrations

A B S T R A C T

The Gaussian Markov Random Field (GMRF) is widely acknowledged for its remarkable flexibility, especially in the realm of high-resolution prediction, when compared to conventional Kriging methods. Rooted in the fundamental principles of Bayesian estimation, this methodology underscores the importance of a meticulous examination of prior and hyperprior distributions, along with their corresponding parameter values. Sensitivity analyses are crucial for evaluating the potential impact of these distributions and parameter values on prediction results. To determine the most effective values for hyperprior parameters, an iterative trial-and-error approach is commonly employed. In our research, we systematically assessed a variety of parameter values through exhaustive cross-validation. Our study is focused on optimizing hyperprior parameter values, with a particular emphasis on Penalized Complexity (PC). We applied our method to conduct spatiotemporal high-resolution predictions of PM2.5 concentrations in Jakarta province, Indonesia. Achieving accurate predictive modeling of PM2.5 concentrations in Jakarta is contingent upon this optimization. We identified that the optimal values for PC hyperprior parameters, with a range of less than 2,000 and a hyperprior standard deviation greater than 1 with a 0.1 probability, yield the most accurate predictions. These parameter values result in the minimum mean absolute percentage error (MAPE) of 19.35393, along with a deviation information criterion (DIC) of -154.23. Our findings highlight that the standard deviation parameter significantly influences model fit compared to the relatively insignificant impact of the range parameter. When coupled with high-resolution mapping, these optimized parameters facilitate a comprehensive understanding of distribution patterns. This process aids in detecting areas particularly susceptible to risks, thereby enhancing decision-making efficacy regarding air quality management.

1. Introduction

Nowadays, there is a growing demand for spatiotemporal high-resolution modeling and mapping, particularly in the field of air pollution analysis (Gladson et al., 2022; Wang et al., 2022). High-resolution mapping provides a comprehensive and detailed perspective on the distribution of air pollution, offering insights into its spatial and temporal variations (Liu et al., 2023). This detailed information is invaluable for guiding government policies aimed at reducing pollution concentrations and mitigating adverse health impacts (Fitriani & Gede Nyoman Mindra Jaya, 2020; Fund Defense, 2020). The process of modeling at high spatiotemporal resolution is intricate and involves numerous factors, leading to substantial computational costs (Jaya & Folmer, 2022). Many techniques are used in high resolution mapping; two commonly used techniques are Inverse Distance Weighting (IDW) and Kriging (Varentsov et al., 2020). IDW boasts simplicity in modeling and rapid computation (Li et al., 2014; Lu & Wong, 2008). However, this system also has significant shortcomings, such as being more sensitive to outliers, tends to underfit, is unable to handle non-stationarity problems, does not take into account spatiotemporal autocorrelation, and is unable to provide information on prediction uncertainty (Lloyd & Atkinson, 2002). To address underfitting concerns,

* Corresponding author.
E-mail address: mindra@unpad.ac.id (I G. N. M. Jaya)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.12.018

the Kriging method emerges as a robust alternative, particularly in dealing with outliers with variogram analysis (Armstrong & Boufassa, 1988; Sun et al., 2019). Kriging considers spatiotemporal autocorrelation and provides predictions with a high level of uncertainty (Handcock, 1994). Despite its effectiveness, the commonly used spatiotemporal Kriging approach has limitations. Kriging struggles with non-stationarity issues and encounters challenges in defining appropriate theoretical semi-variograms when data exhibits high heterogeneity (Liang & Kumar, 2013). Additionally, it fails to account for uncertainty in constructing the semi-variogram and relies on the assumption of normally distributed estimation errors with constant variance, a condition that may not always align with real-world scenarios. The Gaussian Markov Random Field (GMRF) stands out as a more flexible alternative approach (Cai et al., 2013). GMRF introduces a random effect component to enhance the precision of predictions, particularly in high-resolution spatial modeling (Song et al., 2008). Its notable advantages include the ability to address non-stationarity issues, offer flexibility in defining structured spatial dependencies, and account for uncertainty not only in predictions but also in model parameters. This comprehensive approach allows for a more natural interpretation of prediction outputs. GMRF functions within a Bayesian framework, a context not devoid of challenges as emphasized in the literature (Gelman, 2006). One of the key considerations in Bayesian methodology is the specification of prior and hyperprior distributions (Adesina et al., 2018). While Bayesian approaches provide a powerful means to incorporate prior knowledge and update predictions as new data becomes available, the choice and specification of priors can introduce subjectivity and impact model outcomes (Sprenger, 2018). Careful consideration and sensitivity analysis are often necessary to mitigate the influence of prior choices on the results. In conclusion, the GMRF approach, rooted in Bayesian principles, furnishes a comprehensive grasp of uncertainty and enables increased adaptability in spatial modeling, even when confronted with non-stationary conditions (Jaya & Folmer, 2022). However, the thoughtful selection and specification of prior and hyperprior distributions remain critical aspects in ensuring the robustness and reliability of the Bayesian modeling framework (Gómez-Rubio, 2020). There are two general approaches in Bayesian, namely the Markov Chain Monte Carlo (MCMC) and Integrated Nested Laplace Approximation (INLA) methods. Due to the stochastic nature of random effects and the inherent properties of hierarchical models, MCMC convergence is typically both slow and unpredictable. INLA emerged as an alternative method to effectively fit Bayesian with random effects components within the latent Gaussian class, in response to this challenge where significantly reducing the computational time required (Li et al., 2014; Lu & Wong, 2008). Specifying the prior is no longer an issue in the INLA method, as it operates under the assumption that all parameters follow a Gaussian prior distribution. Conversely, establishing hyperpriors for hyperparameters represents the principal obstacle. The formulation of priors requires careful deliberation, as they play a vital role in Bayesian analysis.

Bayesian INLA frequently employs various hyperprior distributions, encompassing log Gamma for log hyperparameter precision, Half-Normal, Half-Cauchy, Half-t, uniform, and Penalized Compliance for hyperparameter standard deviation (Gómez-Rubio, 2020). The last hyperprior distribution is commonly used for reasons to overcome overfitting especially in spatiotemporal high-resolution prediction (Sørbye & Rue, 2017). However, the use of a prior PC can also result in an overly smooth estimate for selected hyperprior parameter values (Ventrucci & Rue, 2016). In high-resolution mapping using Gaussian Markov Random Fields (GMRF), two pivotal parameters demand careful consideration: the range and standard deviation of the Gaussian field. The determination of hyperprior parameter values holds significant importance in ensuring the production of precise and accurate high-resolution predictions (Molina et al., 2008). Evaluating the sensitivity of the hyperprior value for this parameter is essential to attain accurate prediction results (Sørbye & Rue, 2014). In this context, sensitivity analysis typically entails scrutinizing how alterations in input parameters or prior distributions influence the outcomes of Bayesian analysis. So in this study, optimizing PC prior parameter values becomes essential to achieve the most optimal results, particularly in predictive modeling, with the goal of minimizing prediction errors. Various criteria can be employed to assess prior sensitivity and identify the most optimal hyperprior parameter values. Two crucial considerations in this evaluation are fit and predictability. To gauge the model fit, metrics such as Deviance Information Criteria (DIC) and Watanabe-Akaike Information Criteria (WAIC) are utilized. For assessing prediction ability, several measures, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Prediction Error (MAPE), can be examined. The optimal hyperprior parameter values are determined through a surface plot approach, aiming to select values that correspond to the minimum DIC, WAIC, RMSE, MAE, and MAPE (Molina et al., 2008).

Nevertheless, optimization is seldom employed in the estimation of air pollutant concentrations modeling, and there is a lack of justification for the selection of hyperprior parameter values. Meanwhile, accurate predictions of air pollution concentrations are imperative, particularly given the utilization of this information for the development of early warning systems and the prevention of adverse impacts associated with air pollution (Bai et al., 2018). The most dangerous pollutants that have become a major concern are the particles found in pollution (Manisalidis et al., 2020). One very small particulate matter is $PM_{2.5}$ with a diameter of less than 2.5 mm. With its very small size, $PM_{2.5}$ can quickly cause irritation of the respiratory tract. We will apply our optimization process to the PC prior for modeling and mapping $PM_{2.5}$ in Jakarta, the capital of Indonesia. Jakarta has faced significant air pollution challenges. Jakarta has the worst air quality in the world. areas in Jakarta do not have air pollution observation stations due to limited monitoring equipment and expensive costs. The uneven placement of observation stations poses limitations in both the measurement and forecasting of $PM_{2.5}$. (Chen et al., 2023; Nakanishi et al., 2022). Therefore, the monitoring station which only covers an area of 5 km^2 is not representative enough to show air pollution in this area measuring 661.5 km^2 (Nurfaizah, 2022).

The structure of the paper is outlined as follows. Section 2 delineates the Bayesian inference, prediction framework, and INLA. Section 3 details the application of this methodology in producing high-resolution predictions of $PM_{2.5}$ concentrations in Jakarta Province. The ensuing Section 4 serves as the discussion section, while Section 5 encapsulates the conclusion.

2. Methodology

2.1 Spatiotemporal Modeling

Consider $Y(\boldsymbol{s}_i.t)$ as a spatiotemporal process observed at geographical location $\boldsymbol{s}_i$ $(i = 1, \dots, n)$ and time t $(t = 1. \dots . T)$. Let the realization of the spatiotemporal $y(\boldsymbol{s}_i.t)$ be defined by the following equation (Cressie & Wikle, 2012):

$$y(\boldsymbol{s}_i.t) = \mu(\boldsymbol{s}_i.t) + \varepsilon(\boldsymbol{s}_i.t) \tag{1}$$

where $y(\boldsymbol{s}_i.t)$ represents outcomes such as $PM_{2.5}$ at location $\boldsymbol{s}_i$ and time t, while $\mu(\boldsymbol{s}_i.t)$ describes the latent Gaussian spatiotemporal process. The term $\varepsilon(\boldsymbol{s}_i.t)$ represents a measurement error assumed to follow independent and identically distributed (i.i.d) Gaussian distribution with a mean of zero mean and a variance of σ_ε^2 (Lloyd & Atkinson, 2001).

The latent Gaussian spatiotemporal process $\mu(\boldsymbol{s}_i.t)$ is conceptualized as an additive function comprising both temporal and spatiotemporal interaction components.

$$\mu(\boldsymbol{s}_i.t) = \beta_0 + \zeta(t) + \Phi(\boldsymbol{s}_i.t) \tag{2}$$

where β_0 is an intercept that states a fixed effect on all domain regions and time periods, ζ_t is a temporally structured random effect, $\Phi(\boldsymbol{s}_i.t)$ is a Gaussian Field (GF) capturing spatiotemporal autocorrelation. A structured temporal random effect $\zeta(t)$ is defined as a random *walk* process on the first (RW1) or second (RW2) order.

RW1:

$$\varsigma(t+1) - \varsigma(t) | \sigma_\varsigma^2 \sim N\left(0, \sigma_\varsigma^2\right) \text{ for every } i \text{ and } t = 1, \dots, T-1, \tag{3}$$

RW2:

$$\varsigma(t) - 2\varsigma(t+1) + \varsigma(t+2) | \sigma_\varsigma^2 \sim N\left(0, \sigma_\varsigma^2\right) \text{ for every } i \text{ and } t = 1, \dots, T-2, \tag{4}$$

where σ_ς^2 represents the variance hyperparameter of the random walk (RW) process, which governs the smoothness of the underlying process (Rasmussen & Williams, 2005). In this study, we assume that the Gaussian Field (GF) $\Phi(\boldsymbol{s}_i.t)$ undergoes temporal changes according to a first-order autoregressive process (AR1) with a coefficient λ, where $|\lambda| < 1$. The AR1 process characterizes variations in the variable's value concerning its previous values, that is:

$$\Phi(\boldsymbol{s_i}, t) = \lambda \Phi(\boldsymbol{s_i}, t-1) + \gamma(\boldsymbol{s_i}, t) \text{ for t } = 2, \dots, \text{T and i} = 1, \dots, \text{n} \tag{5}$$

with $\Phi(\boldsymbol{s}_i, t) \sim N(0, \frac{\sigma_\gamma^2}{1-\lambda^2})$, characterized by an average of 0 and a variance of $\frac{\sigma_\gamma^2}{1-\lambda^2}$, it is evident that at the initial time $t = 1$, $\Phi(\boldsymbol{s}_i, 1)$ is a random variable sampled from this normal distribution. The term $\gamma(\boldsymbol{s}_i, t)$ represents a mean-square-differentiable process, exhibiting temporal independence or lacking temporal correlation. This process incorporates spatial innovation, which is spatially correlated and drawn from a Gaussian distribution with a mean of 0, along with spatial Matérn covariance functions outlined as follows:

$$\text{Cov}\left(\gamma(\mathbf{s}_i, t), \gamma\left(\mathbf{s}_j, t\right)\right) = \begin{cases} 0, & if\ t \neq t' \\ \sigma_\gamma^2 R(d), & if\ t = t' \end{cases} \tag{6}$$

The parameter σ_γ^2 is a uniform version of $\gamma(\mathbf{s}_i, t)$, and $R(d)$ is a spatial correlation function determined by the distance, Euclidean distance $d = \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \in \mathbb{R}$, between $\mathbf{s}_i$ and $\mathbf{s}_j$. The spatial correlation function is defined as:

$$R(d) = \frac{1}{2^{v-1}\Gamma(v)} (\kappa d)^v K_v(d), \forall t \tag{7}$$

By combining Eq. (6) and Eq. (7), the Matérn covariance function is given by:

$$\sum(d) = \frac{\sigma_\gamma^2}{2^{v-1}\Gamma(v)} (\kappa d)^v K_v(d), \forall t \tag{8}$$

The Matérn covariance function is characterized by three hyperparameters that govern different aspects of the process. The hyperparameter σ_γ^2 governs the marginal variance of the Gaussian Field (GF) process. The hyperparameter κ controls the spatial correlation distance, defined as $r = \frac{\sqrt{8v}}{\kappa}$. For a given value of r, the correlation function will approach 0.13, and the hyperparameter v influences the smoothness of the process. Higher values of v lead to a smoother process. To simplify the

computational process, we introduce the temporal vector GF (Green's Function) denoted as $\boldsymbol{\Phi}_t = (\Phi(\boldsymbol{s}_1, t), \ldots, \Phi(\boldsymbol{s}_n, t))'$, which can now be formally defined as:

$$\boldsymbol{\Phi}_t = \lambda \boldsymbol{\Phi}_{t-1} + \boldsymbol{\gamma}_t \text{ with } \gamma_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \text{ for } t = 2, \ldots, T \tag{9}$$

Continuously Indexed Gaussian Field (GF) typically entails a dense covariance matrix $\boldsymbol{\Sigma}$, posing intricate and time-consuming numerical estimating problems that are sometimes referred to as the '*big n problem*' (Sidén, 2020). The *'big n*' problem emerges when dealing with a substantial number of locations or observation points in spatial studies. To overcome this challenge, Lindgren proposed a solution by substituting the continuously indexed Gaussian Field with a sparse and discretely indexed Gaussian Markov Random Field (GMRF) utilizing Stochastic Partial Differential Equation (SPDE) and Finite Element Method (FEM) approaches (Lindgren et al., 2011; Marhamah & Jaya, 2020). The second-order stationary and isotropic Matérn covariance function's spatially continuous distribution is estimated throughout the research region using a Delaunay triangulation (MESH) in FEM. The fundamental function of the provided triangulation, represented as $\boldsymbol{\gamma}_t$ is determined by the formula that follows:

$$\boldsymbol{\gamma}_t \approx \sum_{l=1}^{L} \psi_l \tilde{\gamma}_{tl} \, \forall t \tag{10}$$

In Eq. (10), where L is the Delaunay triangulation's total number of vertices., $\{\psi_l\}_{l=1}^{L}$ consists of basic functions, and $\tilde{\gamma}_{tl}$ represents the relative weights of all the Gaussian basis functions. $\tilde{\gamma}_{tl}$ is a Gaussian distributed variate with mean 0 and sparse precision matrix $\mathbf{Q}$.

2.2 Penalize Complexity Hyperprior

In Bayesian statistical approach, MCMC is a popular computational tool used to solve various research problems (Jaya et al., 2018). Bayesian methods are flexible and allow researchers to build increasingly complex models. However, as models become more complex, determining the prior distributions or priors becomes more challenging. There are several commonly used options such as prior information, informative priors, or non-subjective priors (Gelman et al., 2013). R-INLA device developers can either force users to specify all prior shared priors explicitly or provide default priors (Rue et al., 2009; Martins et al., 2013). Informative PC priors follow four basic principles and offer flexibility in changing those principles as needed. These PC priors can help build better and more transparent priors.

Prior Penalized Complexity (PC) is a practical method to construct the prior in Bayesian statistics. Prior PCs impose penalties against deviations from the base model and encourage simplicity until there is enough evidence for more complex models (Fuglstad et al., 2019). It is based on Occam's razor principle, which prioritizes simpler explanations until there is support for more complex explanations. This helps prevent overfitting and improves the model's ability to generalize and predict (Simpson et al., 2017). The PC prior, in addition to its application of constant rate penalization, embraces the concept of penalizing deviations from the base model at a uniform rate. Furthermore, the PC prior offers the flexibility for users to customize scaling, enabling them to regulate the degree of flexibility granted to the model. This scaling parameter can be adjusted based on weak information or other subjective criteria, empowering users to tailor the model to their specific needs (Simpson et al., 2017). Prior PCs provide flexibility in determining prior on parameters that are difficult to determine directly from expert knowledge. They can be vague, weakly informative, or strongly informative according to the tuning parameters chosen by the researcher (Evans & Jang, 2011; Gelman et al., 2013).

The PC prior concerning the precision parameter τ in the context of a Gaussian random effect is established through an evaluation of the natural base model, where τ is set to ∞. The PC prior introduces a penalty that quantifies the increase in complexity compared to the base model when utilizing a more flexible model (Bhattacharya et al., 2014). This penalty decreases gradually as the complexity increases, ensuring that simpler models are preferred unless there is substantial evidence in favour of a more intricate one (Lindgren & Rue, 2015). The PC prior for the precision of a Gaussian random effect proves to be a valuable tool in modeling hierarchical structures, as it effectively captures relationships and unobserved variations. It accommodates the incorporation of random effects that enhance model flexibility while preserving a preference for simplicity, striking a balance between adaptability and parsimony (Simpson et al., 2017).

In this study, To complete this model, we need to define hyperprior distributions for several parameters, such as range (r), spatial autoregression parameter (ρ), temporal autocorrelation parameter (λ), and standard deviation (σ) for various components such as $\beta_0, \zeta, \Phi, \gamma$ and ε. Typically, we give a large value, for example $\sigma_{(\beta_0)} = 10^6$, for the standard deviation of the hyperparameter $\sigma_{(\beta_0)}$. This allows the intercept to have a large range of potential values, increasing the flexibility of the model. For the other parameters, we use hyperprior complexity dialysis (PC) (Simpson et al., 2017). Hyperprior is commonly used to prevent overfitting problems. $Prob(\sigma > \mu_\sigma) = \alpha_\sigma$ is used to define a hyperprior PC for standard deviation, where $0 < \alpha_\sigma < 1$ is the probability value and $\mu_\sigma > 0$ is a hyperprior quantile. For all $\sigma_\zeta, \sigma_\Phi, \sigma_\lambda, \sigma_\rho$, and σ_ε.

2.3 Sensitivity

To assess the sensitivity analysis and ascertain the most effective values for the PC hyperprior parameters, six combinations of parameter ranges (r) and standard deviations (σ_γ) were investigated. We use two different criteria to evaluate the

sensitivity and find the optimal prior. The primary criteria include assessing the model fit using the Deviance Information Criterion (DIC) and the Watanabe-Akaike Information Criterion (WAIC) (Blangiardo, 2015; Fassò et al., 2007). DIC is defined as:

$$\mathrm{DIC} = 2\overline{\mathrm{D}} - \mathrm{D}(\overline{\boldsymbol{\theta}}) \tag{11}$$

where $\overline{\mathrm{D}} = \mathrm{E}_{\theta|\mathrm{y}}\big(\mathrm{D}(\boldsymbol{\theta})\big)$ is the posterior mean of the deviance value $\mathrm{D}(\boldsymbol{\theta})$ and is the deviance value $\mathrm{D}(\boldsymbol{\theta}) = -2\log\big(\mathrm{p}(\mathbf{y}|\boldsymbol{\theta})\big)\mathrm{D}(\overline{\boldsymbol{\theta}})$ of the posterior mean of the parameter. WAIC can be formulated as:

$$\mathrm{WAIC} = -2(\mathrm{D} - \mathrm{p}_{\mathrm{WAIC}}) \tag{12}$$

where D denotes the deviance measuring the model fit, i.e. $\mathrm{D} = \sum_{i=1}^{n}\sum_{t=1}^{T}\log E_{\theta|\mathrm{y}}[p(y_{it}|\boldsymbol{\theta})]$ and $\mathrm{p}_{\mathrm{WAIC}}$ denotes the effective number of parameters, i.e., $\mathrm{p}_{\mathrm{WAIC}} = \sum_{i=1}^{n'}\sum_{t=1}^{T}\mathrm{Var}_{\mathrm{posterior}}[\log p(y_{it}|\boldsymbol{\theta})]$. The smaller DIC and WAIC value indicate a better fit. Additionally, a secondary set of criteria focuses on evaluating the model's predictive capabilities, which comprises the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Prediction Error (MAPE). The three criteria for model predictive performance were calculated using the following formulas:

$$MAE = \frac{\sum_{i=1}^{n'}\sum_{t=1}^{T}\left|y_{obs,i} - y_{pre,i}\right|}{n'T} \tag{13}$$

$$RMSE = \frac{\sum_{i=1}^{n'}\sum_{t=1}^{T}\left(y_{obs,i} - y_{pre,i}\right)^2}{n'T} \tag{14}$$

$$MAPE = \frac{\sum_{i=1}^{n'}\sum_{t=1}^{T}\left|\frac{y_{obs,i} - y_{pre,i}}{y_{pre,i}}\right|}{n'T} \times 100 \tag{15}$$

In Equations (13-15), $y_{obs,i}$ and $y_{val,i}$ represents the testing observation i and the testing prediction i with n' denotes the number of testing units. The smaller MAE, RMSE, and MAPE value indicate a better predictive performance. We will select the hyperprior parameter values based on a surface plot that offers the optimal combination of both fit and predictive performance.

3. Application: High-resolution prediction on $PM_{2.5}$ in Jakarta Province, Indonesia

3.1 Exploratory Analysis of $PM_{2.5}$ Observation

We performed a descriptive analysis of $PM_{2.5}$ concentration at 8 observation stations in Jakarta province and the finding are detailed in Table 1. The results indicate that the minimum concentration of $PM_{2.5}$, recorded at 21.48 μg/m^3, was observed in February 2022, while the maximum concentration, reaching 68,589 μg/m^3, occurred in September 2022. On average, the concentration of $PM_{2.5}$ throughout 2022 was approximately 36.96 μg/m^3. Furthermore, we analyzed fluctuations in $PM_{2.5}$ concentrations by examining the standard deviation values. It is evident that during June, July, and August, the standard deviation values are notably higher compared to other months. This observation indicates substantial fluctuations in $PM_{2.5}$ concentrations during this three-month period.

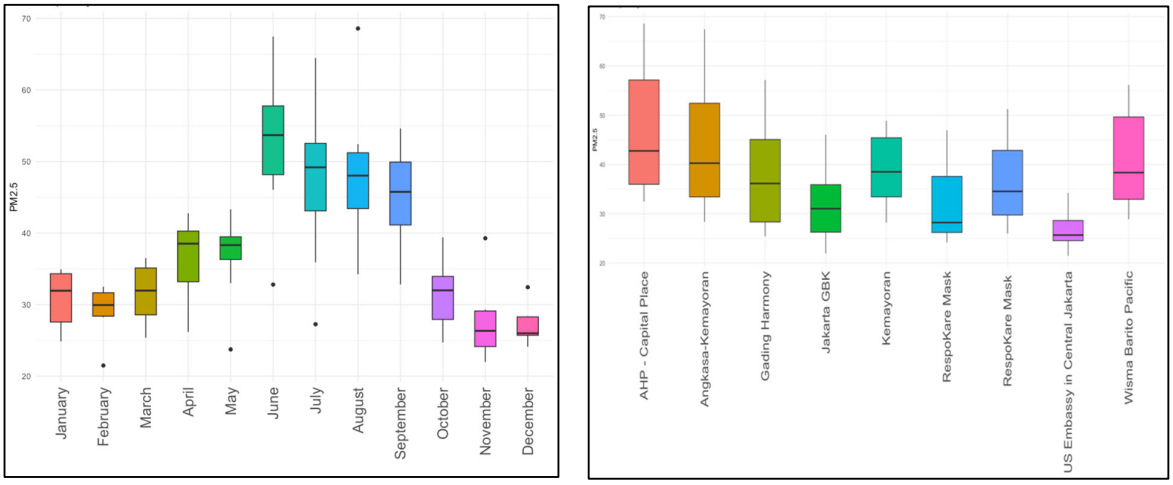
Table 1
Statistical characteristics of the $PM_{2.5}$ concentration observations (μg/m^3)

Time	Mean	Median	SD	Min	Max
January	30.813	31.939	4.316	24.854	34.932
February	28.994	29.937	4.042	21.486	32.502
March	31.600	31.957	4.486	25.376	36.505
April	36.276	38.519	5.861	26.177	42.766
May	36.771	38.316	6.050	23.735	43.322
June	52.421	53.687	10.382	32.800	67.463
July	47.212	49.190	11.437	27.259	64.458
August	48.142	48.023	10.454	34.236	68.590
September	44.864	45.761	7.626	32.814	54.621
October	31.485	31.990	4.866	24.714	39.397
November	27.723	26.333	5.743	21.978	39.270
December	27.222	25.993	2.747	24.121	32.434

Note: SD represents the standard deviation, Min indicates to the minimum values, and Max indicates the maximum values of PM2.5 concentrations.

To identify potential anomalies in the data, we have included a boxplot in Fig. 1. Fig. 1(a) illustrates the distribution of $PM_{2.5}$ concentrations in the temporal domain, while Fig. 1(b) focuses on the spatial domain. Fig. 1(a) shows several outliers, indicating extreme values that suggest unusual events influencing $PM_{2.5}$ concentrations in specific months (i.e., August). The boxplot also shows a skewed distribution, signifying a specific pattern in the temporal variation. This suggests a tendency for higher concentrations of $PM_{2.5}$ in certain months. Fig. 1(b) shows there is no outlier. However, all stations exhibit skewness in the spatial variation of $PM_{2.5}$ concentrations. This indicates a distinct pattern in the spatial distribution of $PM_{2.5}$

concentrations across the monitoring stations, even in the absence of outliers. This condition has the potential to impact the accuracy of prediction results significantly. Hence, in this study, log transformations were applied to the data. The aim of this transformation is to maintain the stability of data variance, make the data distribution more symmetrical, and reduce the impact of outliers. Through log transformations, extreme values are expected to exert less impact, addressing distribution irregularities and enhancing the performance of predictive analytics.



(a) Temporal Domain (b) Spatial Domain

Fig. 1. Boxplot data of $PM_{2.5}$ concentrations

Analysis of Fig. 1 reveals distinctive spatial patterns in $PM_{2.5}$ concentrations across the 8 observation stations from January to December 2022. Significant variations are observed, particularly with higher concentrations around the AHP and Angkasa Kemayoran observation stations situated in Central Jakarta, known as an industrial hub. This industrial activity is identified as a likely primary contributor to the elevated $PM_{2.5}$ levels in this area. In contrast, the US Embassy observation station recorded the lowest $PM_{2.5}$ concentration. This area experiences comparatively lower pollution levels due to factors like limited industrial activities, reduced vehicular traffic, and fewer human activities. The embassy area is not a commonly traversed public route, emphasizing the role of gas emissions from motor vehicles as a major pollution contributor. Furthermore, the embassy area benefits from a strong system for monitoring air quality, allowing for early detection and effective control of air pollution, thereby ensuring that emission levels are well-regulated.

3.2 Spatiotemporal Modeling and Estimations of the PM2.5 Concentrations

3.2.1 Parameter Construction

In the modeling process, we performed cross-validation by randomly selecting 3 out of the 8 monitoring stations (37.5%) as validation data. To assess the sensitivity of the parameters to the estimated results, we tested six combinations of parameter ranges (in km) r = {2, 4, 6, 10, 13, 20} and standard deviations σ_γ = {0.01, 0.05, 0.1, 0.5, 1, 1.5}. Following a systematic exploration of various values, we evaluated the model's goodness using DIC to assess the prediction model's suitability and accuracy. Model accuracy was measured through MAPE instead of RMSE and MAE due to their negligible differences. The results indicated optimal parameters with a range parameter value of 2 km and a standard deviation set to 1 (MAPE = 19.35393, DIC = -154.23). Full results can be found in Table 2. Table 2 displays an average spatial range of the field at 2,792 km, suggesting that the spatial dependence is estimated to extend up to 2,792 km. Additionally, a rho parameter of 0.862 indicates significant temporal variation in the estimated $PM_{2.5}$ concentrations.

Table 2
Hyperparameter Result

	Mean	SD	$q_{0.025}$	$q_{0.5}$	$q_{0.975}$
Precision for the Gaussian observations	497.287	600.821	47.224	316.579	2054.712
Precision for Time	34.215	16.439	11.965	31.002	75.077
Range (r) for spatial field	2792.003	1763.017	765.775	2358.299	7394.089
SD (σ_γ) for spatial field	0.161	0.049	0.087	0.154	0.278
Group Rho for spatial field	0.862	0.191	0.286	0.932	0.996

Note: SD is Standard Deviation, $q_{0.025}$ and $q_{0.975}$ represent the credible interval, $q_{0.025}$ is the lower quantile or the 2.5th quantile, $q_{0.5}$ is the median value of the range distribution, and $q_{0.975}$ is the upper quantile or the 97.5th quantile.

3.2.2 Result Verification

Utilizing parameter estimation, $PM_{2.5}$ estimation modeling through Gaussian Markov Random Fields (GMRF) was conducted on a 34 × 34 km grid. The intention was to identify the parameter with a range of 2 km, considering 5 different standard deviation parameter values, as the optimal choice based on the smallest Deviance Information Criterion (DIC) and Mean Absolute Percentage Error (MAPE), thus ensuring the consistency of the model.

Table 3
Five Smallest MAPE and DIC

Range (km)(r)	σ_γ	DIC	WAIC	R	RMSE	MAPE
2	1	-154.23	-130.70	0.693	2.001	19.354
2	1.5	-152.64	-129.93	0.693	2.006	19.369
2	0.5	-147.91	-129.00	0.692	1.954	19.391
2	0.1	-121.44	-115.36	0.685	1.924	19.555
2	0.05	-114.56	-110.67	0.683	1.890	19.577

Note: σ_γ is standard deviation parameter, R is correlation.

We generated a 3D surface plot to visually assess different combinations of range and standard deviation values. This plot allowed us to examine the relationship between the response variable and two predictor variables in a three-dimensional space, providing insights into preferred response values and operational scenarios. Peaks and valleys on the plot correspond to combinations of x and y generating local maxima or minima. In Fig. 2, the contour plot illustrates the connection between various scenarios of parameter range values and standard deviation parameter values used for predicting $PM_{2.5}$ concentrations and evaluating the performance metrics of the resulting model (DIC and MAPE). A narrower range with a higher standard deviation parameter is associated with smaller prediction errors. The optimal result, indicated by the deep valley in the plot, aligns with the best model evaluation, achieved with a range parameter set to 2 km and a standard deviation parameter set to 1.

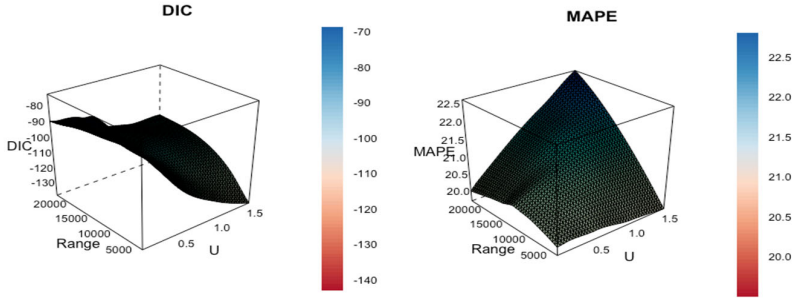


Fig. 2. 3D surface plot of predictions $PM_{2.5}$ Concentration based on evaluation metric DIC and MAPE

In Table 3, it is evident that the correlation is 0.693 for the parameter range set to 2 km and the standard deviation of the parameter set to 1. This signifies a robust relationship between the predicted and observed results, with a value closer to 1 indicating a stronger association between the two variables. The substantial correlation suggests that the predictions effectively explain variations in $PM_{2.5}$ concentrations, implying that the model utilized in the analysis possesses reliable predictive capabilities. Besides the correlation value, we evaluate the model's performance using scatter plots that compare observed results with predicted results at three validation monitoring locations.

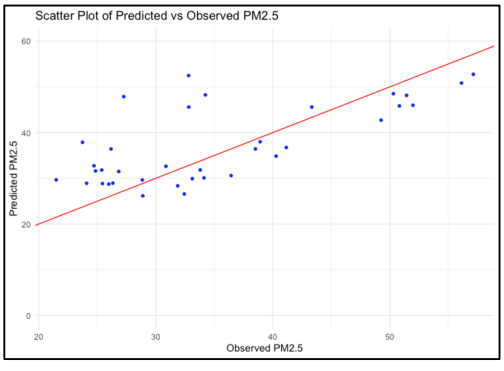


Fig. 3. Scatterplots of observed $PM_{2.5}$ concentrations and cross-validation predictions

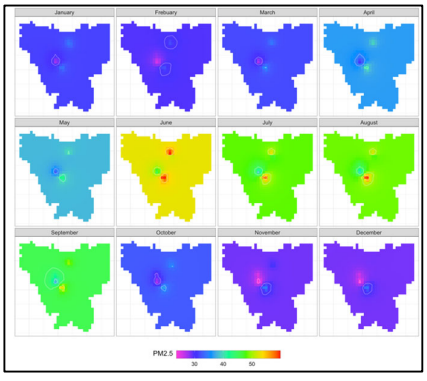


Fig. 4. Variability in spatiotemporal of the predicted $PM_{2.5}$ concentrations in DKI Jakarta

The results presented in Fig. 3 reveal that most data points closely align with the reference line, indicating strong performance in estimating overall $PM_{2.5}$ concentrations. The results presented in Fig. 4 reveal an interesting phenomenon. The highest concentrations of $PM_{2.5}$ are observed in June, July, and August, as depicted by the color map. The red dots in the central and northern Jakarta areas signify significantly higher concentrations of $PM_{2.5.}$ This occurrence is likely due to these three months being part of the dry season in the region. During the dry season, low rainfall and dry weather conditions allow dust and small particles, such as $PM_{2.5}$ concentrations, to remain unaffected by rain and more easily transported by wind. The model shows increased sensitivity to higher $PM_{2.5}$ concentrations, accurately reflecting their distribution in heavily polluted areas. $PM_{2.5}$ concentrations range from 23.758 μg/m³ as the minimum value to 59.830 μg/m³ as the maximum value.

Simpson et al. (2017) introduced the concept of the Penalized Complexity (PC) hyperprior, which has demonstrated effectiveness of spatiotemporal predictions using GMRF modeling. The PC prior is computed based on specific principles that evaluate the complexity of model components by assessing their deviation from simple base model formulations. This approach has found widespread application in high-resolution spatiotemporal predictions, particularly where GMRF modeling is prevalent. As emphasized by Simpson et al. (2017), the PC prior plays a pivotal role in mitigating the risk of overfitting in prediction models. However, it's crucial to acknowledge that the hyperprior PC exhibits high sensitivity in determining parameter values. In GMRF modeling, precise parameter values are paramount for accurate spatiotemporal predictions, particularly in high-risk scenarios. Two such parameters requiring careful prior definition are the range parameter and the standard deviation. In practice, identifying optimal values for these parameters can be challenging, often necessitating an iterative trial-and-error approach. In our study focused on predicting $PM_{2.5}$ concentrations in Jakarta, we aimed to identify the most suitable parameter values by systematically testing various values and conducting cross-validation. We considered the suitability of the prediction model based on DIC values and the accuracy of the model in providing predictions, measured through MAE, RMSE, and MAPE. For cross-validation, we divided the data into training and testing datasets, allocating five data points to the former and three data points to the latter. To pinpoint the most optimal parameter values, we constructed a 3D surface plot, allowing us to visualize and evaluate different combinations of range and standard deviation values. Our findings revealed that, for our specific case, the most optimal parameter values were a range of 2 km and a standard deviation of 1. Notably, we observed that for prediction purposes, a higher range value was associated with a smaller standard deviation value, and vice versa, to yield the most accurate predictions. On the other hand, concerning the model fit, it became evident that the range parameter had a relatively minor impact, with the standard deviation parameter being the most dominant and influential factor.

4. Conclusion

This study utilized GMRF spatiotemporal modeling, with a focus on selecting appropriate hyperprior parameter values for hyperparameters in Bayesian setting to achieve accurate prediction. Specifically, the study employed the Penalized Complexity (PC) hyperprior that is usually used to handle the overfitting issues in complex models. The optimization of the PC hyperprior parameter values were a crucial aspect of this research, aiming to obtain the most optimal results tailored to the requirements, especially for predictive purposes, with the goal of minimizing prediction errors. The optimization of PC hyperpriors parameter values will be applied for spatiotemporal high-resolution modeling and mapping PM2.5 in Jakarta. This strategic approach aims to fine-tune the model to meet specific requirements and enhance its predictive accuracy, contributing valuable insights to air quality management and decision-making processes.

This analytical technique aimed to assess how variations in hyperprior parameters, particularly within the Penalized Complexity (PC) framework, impacted the Spatiotemporal GMRF modeling's performance in estimating $PM_{2.5}$ concentrations. Two critical parameters, the range parameter (r) and standard deviation (σ_y) were identified as pivotal contributors to the model's effectiveness. The study explored a range of scenarios, varying the range parameter (in km) r = {2, 4, 6, 10, 13, 20} and adjusting the standard deviation σ_γ = {0.01, 0.05, 0.1, 0.5, 1, 1.5}. This comprehensive approach facilitated an in-depth understanding of the model's behavior under different parameter configurations. To refine the optimization process, the study delved into simplified optimization approaches and compared various cross-validation techniques. The assessment, based on MAPE and DIC criteria, pinpointed the optimal parameters for $PM_{2.5}$ concentration estimation using GMRF, determining a distance of 2 km and a standard deviation of 1 as the most effective combination.

Innovative visualization methods, extending beyond traditional 3D surface plots, were employed to unravel intricate parameter interactions. These advanced techniques provided nuanced insights into the spatial-temporal dynamics of $PM_{2.5}$ distribution in Jakarta. The study's results not only delivered accurate high-resolution predictions but also furnished valuable tools for mapping visualization, aiding in the comprehension of $PM_{2.5}$ distribution patterns.

Based on the discussion and conclusion, the following recommendations are suggested for future studies, firstly, it would be beneficial to fine-tune the Penalized Complexity (PC) hyperprior. Adjusting its details could enhance its effectiveness, especially in situations where accurate predictions in high-risk scenarios are crucial. Furthermore, we propose a thorough exploration of potential modifications to the PC hyperprior. This involves examining different adjustments to strengthen its adaptability and robustness in dealing with a variety of predictive challenges. Adapting the hyperprior to specific scenarios can optimize its usefulness across a range of conditions.

In addition, we emphasize the importance of a detailed sensitivity analysis of the hyperprior's parameters, focusing particularly on the range and standard deviation. This careful examination is expected to provide valuable insights into how these parameters dynamically interact and influence the model's results.

On the other hand, this study helps management take corrective actions to pay more attention to areas estimated to have high levels of $PM_{2.5}$ concentrations. Second, in certain months which are dry seasons with high concentrations of $PM_{2.5}$, people can take preventive measures such as avoiding outdoor activities or using air masks. Third, these predictions enable faster decision-making in sectors such as transportation, energy, and industry.

References

Adesina, O. S., Olatayo, T. O., Agboola, O. O., & Oguntunde, P. E. (2018). Bayesian Dirichlet process mixture prior for count data. *International Journal of Mechanical Engineering and Technology*, *9*(12), 630–646.

Armstrong, M., & Boufassa, A. (1988). Comparing the robustness of ordinary kriging and lognormal kriging: Outlier resistance. *Mathematical Geology*, *20*(4), 447–457. https://doi.org/10.1007/BF00892988

Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air Pollution Forecasts: An Overview. *International Journal of Environmental Research and Public Health*, *15*(4), 780. https://doi.org/10.3390/ijerph15040780

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2014). *Dirichlet-Laplace priors for optimal shrinkage*. https://doi.org/https://doi.org/10.48550/arXiv.1401.5398

Blangiardo, M. (2015). Spatial modeling. *Spatial and Spatio-Temporal Bayesian Models with R - INLA*, 173–234. https://doi.org/10.1002/9781118950203.ch6

Cai, Z., Jermaine, C., Vagena, Z., Logothetis, D., & Perez, L. L. (2013). The Pairwise Gaussian Random Field for High-Dimensional Data Imputation. *2013 IEEE 13th International Conference on Data Mining*, 61–70. https://doi.org/10.1109/ICDM.2013.149

Chen, J., Miao, C., Yang, D., Liu, Y., Zhang, H., & Dong, G. (2023). *Estimation of fine-resolution PM 2 . 5 concentrations using the INLA-SPDE method*. *14*(April).

Cressie, N., & Wikle, C. K. (2012). Space-Time Kalman Filter. In *Encyclopedia of Environmetrics*. Wiley. https://doi.org/10.1002/9780470057339.vas037

Evans, M., & Jang, G. H. (2011). Weak Informativity and the Information in One Prior Relative to Another. *Statistical Science*, *26*(3). https://doi.org/10.1214/11-STS357

Fassò, A., Cameletti, M., & Fass, A. (2007). *Web Working Papers by The Italian Group of Environmental Statistics A general spatio-temporal model for environmental data A general spatio-temporal model for environmental data*. *February*.

Fitriani, R., & Gede Nyoman Mindra Jaya, I. (2020). Spatial modeling of confirmed COVID-19 pandemic in East Java Province by geographically weighted negative binomial regression. *Communications in Mathematical Biology and Neuroscience*, *2020*, 1–17. https://doi.org/10.28919/cmbn/4874

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2019). Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *Journal of the American Statistical Association*, *114*(525), 445–452. https://doi.org/10.1080/01621459.2017.1415907

Fund Defense, E. (2020). *Policies to Reduce Pollution and protect Health*.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3). https://doi.org/10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. https://doi.org/10.1201/b16018

Gladson, L., Garcia, N., Bi, J., Liu, Y., Lee, H. J., & Cromar, K. (2022). Evaluating the Utility of High-Resolution Spatiotemporal Air Pollution Data in Estimating Local PM2.5 Exposures in California from 2015–2018. *Atmosphere*, *13*(1), 85. https://doi.org/10.3390/atmos13010085

Gómez-Rubio, V. (2020). *Bayesian Inference with INLA*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315175584

Handcock, M. S. (1994). *Measuring the Uncertainty in Kriging* (pp. 436–447). https://doi.org/10.1007/978-94-011-0824-9_46

Jaya, I. G. N. M., & Folmer, H. (2022). Spatiotemporal high-resolution prediction and mapping: methodology and application to dengue disease. *Journal of Geographical Systems*, *24*(4), 527–581. https://doi.org/10.1007/s10109-021-00368-0

Jaya, I. G. N. M., Ruchjana, B. N., Tantular, B., Zulhanif, & Andriyana, Y. (2018). Simulation and application of the spatial autoregressive geographically weighted regression model (SAR-GWR). *ARPN Journal of Engineering and Applied Sciences*, *13*(1), 377–385.

Li, L., Losser, T., Yorke, C., & Piltner, R. (2014). Fast Inverse Distance Weighting-Based Spatiotemporal Interpolation: A Web-Based Application of Interpolating Daily Fine Particulate Matter PM2.5 in the Contiguous U.S. Using Parallel Programming and k-d Tree. *International Journal of Environmental Research and Public Health*, *11*(9), 9101–9141. https://doi.org/10.3390/ijerph110909101

Liang, D., & Kumar, N. (2013). Time-space Kriging to address the spatiotemporal misalignment in the large datasets. *Atmospheric Environment*, *72*, 60–69. https://doi.org/10.1016/j.atmosenv.2013.02.034

Lindgren, F., & Rue, H. (2015). Bayesian Spatial Modelling with R - INLA. *Journal of Statistical Software*, *63*(19). https://doi.org/10.18637/jss.v063.i19

Lindgren, F., Rue, H., & Lindström, J. (2011). An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society Series B: Statistical*

Methodology, *73*(4), 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x
Liu, J., Wan, G., Liu, W., Li, C., Peng, S., & Xie, Z. (2023). High-dimensional spatiotemporal visual analysis of the air quality in China. *Scientific Reports*, *13*(1), 5462. https://doi.org/10.1038/s41598-023-31645-1
Lloyd, C., & Atkinson, P. (2001). Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers & Geosciences*, *27*(8), 929–937. https://doi.org/10.1016/S0098-3004(00)00132-1
Lloyd, C. D., & Atkinson, P. M. (2002). Non-stationary Approaches for Mapping Terrain and Assessing Prediction Uncertainty. *Transactions in GIS*, *6*(1), 17–30. https://doi.org/10.1111/1467-9671.00092
Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*, *34*(9), 1044–1055. https://doi.org/10.1016/j.cageo.2007.07.010
Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, *8*. https://doi.org/10.3389/fpubh.2020.00014
Marhamah, E., & Jaya, I. G. N. M. (2020). Modeling positive COVID-19 cases in Bandung City by means geographically weighted regression. *Communications in Mathematical Biology and Neuroscience*. https://doi.org/10.28919/cmbn/4991
Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, *67*, 68–83. https://doi.org/10.1016/j.csda.2013.04.014
Molina, R., Vega, M., Mateos, J., & Katsaggelos, A. K. (2008). Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images. *Applied and Computational Harmonic Analysis*, *24*(2), 251–267. https://doi.org/10.1016/j.acha.2007.03.006
Nakanishi, Y., Kaneta, T., & Nishino, S. (2022). A Review of Monitoring Construction Equipment in Support of Construction Project Management. *Frontiers in Built Environment*, *7*. https://doi.org/10.3389/fbuil.2021.632593
Nurfaizah, A. (2022). *Stasiun Pemantauan Kualitas Udara Jakarta Masih Minim*. Kompas. https://www.kompas.id/baca/metro/2022/12/10/jakarta-perlu-tambah-alat-pemantaun-kualitas-udara
Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning*. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001
Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x
Sidén, P. (2020). *Scalable Bayesian spatial analysis with Gaussian Markov random fields* (Issue Dissertation). www.liu.se
Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, *32*(1), 1–28. https://doi.org/10.1214/16-STS576
Song, H.-R., Fuentes, M., & Ghosh, S. (2008). A comparative study of Gaussian geostatistical models and Gaussian Markov random field models1. *Journal of Multivariate Analysis*, *99*(8), 1681–1697. https://doi.org/10.1016/j.jmva.2008.01.012
Sørbye, S. H., & Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, *8*, 39–51. https://doi.org/10.1016/j.spasta.2013.06.004
Sørbye, S. H., & Rue, H. (2017). Penalised Complexity Priors for Stationary Autoregressive Processes. *Journal of Time Series Analysis*, *38*(6), 923–935. https://doi.org/10.1111/jtsa.12242
Sprenger, J. (2018). The objectivity of Subjective Bayesianism. *European Journal for Philosophy of Science*, *8*(3), 539–558. https://doi.org/10.1007/s13194-018-0200-1
Sun, X.-L., Wu, Y.-J., Zhang, C., & Wang, H.-L. (2019). Performance of median kriging with robust estimators of the variogram in outlier identification and spatial prediction for soil pollution at a field scale. *Science of The Total Environment*, *666*, 902–914. https://doi.org/10.1016/j.scitotenv.2019.02.231
Varentsov, M., Esau, I., & Wolf, T. (2020). High-Resolution Temperature Mapping by Geostatistical Kriging with External Drift from Large-Eddy Simulations. *Monthly Weather Review*, *148*(3), 1029–1048. https://doi.org/10.1175/MWR-D-19-0196.1
Ventrucci, M., & Rue, H. (2016). Penalized complexity priors for degrees of freedom in Bayesian P-splines. *Statistical Modelling*, *16*(6), 429–453. https://doi.org/10.1177/1471082X16659154
Wang, Y., Huang, C., Hu, J., & Wang, M. (2022). Development of high-resolution spatio-temporal models for ambient air pollution in a metropolitan area of China from 2013 to 2019. *Chemosphere*, *291*, 132918. https://doi.org/10.1016/j.chemosphere.2021.132918