JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 9, No. 2, April 2025, pp. 568-586

Spatio-Temporal Median Polish Kriging with ARIMA Integration for Monthly Precipitation Interpolation in East Kalimantan

Friendtika Miftaqul Jannah[1], Rahma Fitriani[1], Henny Pramoedyo[1]
[1]Department of Statistics, Brawijaya University, Indonesia
friendtika@student.ub.ac.id

Article History:
Received : 30-01-2025
Revised : 22-03-2025
Accepted : 24-03-2025
Online : 29-04-2025

Keywords:
ARIMA;
Kriging;
Median Polish;
Precipitation;
Spatio-Temporal.

ABSTRACT

Precipitation can lead to disasters like droughts and floods, necessitating accurate interpolation methods. Traditional spatio-temporal kriging often struggles with outliers, which can reduce estimation reliability. This study develops spatio-temporal median polish kriging, which separates spatial and temporal components to improve interpolation accuracy, particularly in handling outliers. Unlike conventional kriging, this method integrates median polish kriging for robust spatial interpolation and ARIMA for capturing temporal trends, making it more effective in dynamic precipitation pattern estimation. The study utilizes precipitation data from seven observation posts in East Kalimantan (2021–2023). The data is processed using a combination of spatial, temporal, and spatial-temporal modeling approaches to capture precipitation variations accurately. For spatial interpolation, the study applies kriging in median polish spatial effects. The best semivariogram model for spatial effects is exponential, which is used to characterize spatial dependencies. To capture temporal effects of median polish, the study employs ARIMA(1,2,0), which models precipitation trends over time and helps manage temporal fluctuations. For residuals of median polish interpolation, the study applies spatio-temporal kriging, using a simple sum-metric model as the best approach to integrate both spatial and temporal dependencies. The semivariograms selected for spatial, temporal, and joint dependencies follow a gaussian structure. The interpolation results reveal that precipitation increases toward the west, with precipitation patterns also showing an increasing trend over time. These findings demonstrate the model's capability in capturing spatial and temporal precipitation variations while addressing potential outliers through the median polish approach. By utilizing a robust statistical framework, the model reduces the influence of extreme values, leading to more reliable precipitation estimates. However, this study utilizes only seven observation posts. The limited number of observation posts may introduce uncertainty in regions distant from measurement stations and affect the model's accuracy. Therefore, further research should test this model by applying it to different geographical regions with a more extensive dataset.

https://doi.org/10.31764/jtam.v9i2.29570

A. INTRODUCTION

Precipitation plays an essential role in the hydrological cycle, influencing the likelihood of floods and droughts in each area (Katipoğlu, 2022). Meteorological and climatological events, such as storms, atmospheric boundaries (air mass transitions), and local geography, can lead to substantial variations in precipitation over relatively small areas (Adiguna et al., 2021; Pertiwi et al., 2015). As a result, precipitation exhibits significant spatial and temporal

variability, often including extreme values or outliers that complicate accurate prediction and modeling (Zhang et al., 2020).

East Kalimantan, as the province where the National Capital City (IKN) is being relocated, is undergoing extensive environmental changes due to rapid urbanization. The transformation of forests and open land into impervious surfaces alters hydrological processes, leading to increased surface runoff, reduced groundwater recharge, and heightened flood risks during extreme precipitation events (Yang et al., 2020). This is in line with research by Han et al. (2022) and Sari & Atsidiqi (2020), which states that the process of urbanization changes the characteristics of the earth's surface, contributing to phenomena such as the urban heat island (UHI), which can trigger changes in weather patterns and increase the likelihood of precipitation in urban areas. Additionally, deforestation reduces soil water retention capacity, further exacerbating flooding during heavy precipitation and increasing drought susceptibility during dry seasons (Sudinda, 2020). These environmental changes highlight the urgent need for a reliable precipitation interpolation method that can accommodate spatio-temporal variability and extreme values.

Space-time kriging methods have been widely used for spatio-temporal interpolation; however, they are sensitive to outliers and often fail to provide robust predictions when dealing with highly variable precipitation data (Hengl et al., 2007). To address this limitation, Median Polish Kriging (MPK) has been introduced as an alternative approach that effectively handles outliers by decomposing data into trend and residual components with an additive structure by iteratively calculating the median values of rows and columns (Tutmez, 2014; Sun & Genton, 2012). The median polish kriging method improves prediction accuracy by applying median polish, which reduces the influence of extreme values, followed by kriging interpolation on the residuals (Barbara & Wu, 2003).

The research conducted by Tutmez (2014) demonstrated that the two-dimensional (longitude and latitude) median polish kriging can handle outliers by utilizing the median component. Martínez et al. (2017) extended median polish kriging to a four-dimensional spatio-temporal framework (longitude, latitude, altitude, and time) by integrating the ARIMA model for temporal trends, achieving an improved precipitation interpolation with an R^2 value of 0.75. However, Martínez et al. (2017) modeled spatial effects separately, resulting in three distinct spatial effects. This approach complicated simultaneous predictions of spatial effects, as it required additional manual interpolation steps.

To address these limitations, this study focuses on simultaneously modeling spatial effects to simplify interpolation. The proposed method, spatio-temporal median polish kriging (ST-MPK), integrates kriging for simultaneous spatial effects and ARIMA for temporal effects within the median polish framework. Unlike conventional space-time kriging, which is highly sensitive to outliers, ST-MPK enhances robustness by systematically decomposing spatial and temporal trends before interpolation, reducing the influence of extreme values. Additionally, unlike the approach of Martínez et al. (2017), which treated spatial effects separately, this study enables a unified spatial modeling process,

By overcoming these challenges, spatio-temporal median polish kriging with ARIMA integration offers a more accurate and resilient precipitation interpolation method, which is crucial for enhancing water conservation strategies and flood mitigation efforts in East

Kalimantan. Furthermore, this study advances spatio-temporal interpolation methodologies by providing a framework that can be adapted to other regions and climate variables, thereby supporting future research in hydroclimatology and environmental modeling.

B. METHODS

This study relies on secondary data sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG) in Samarinda. The data includes longitude, latitude, and total monthly precipitation collected from seven observation points: Temindung Station (Samarinda), Sultan Aji Station (Balikpapan), Kalimarau Station (Berau), Sangkulirang Post (East Kutai), Kembang Jenggut Post (Kutai Kartanegara), Long Iram (West Kutai), and Gunung Telihan (Bontang) over the period from January 2021 to December 2023. The study employs a spatio-temporal median polish kriging, carried out through the following analytical procedures.

1. Outliers Checking

Outliers are data points that deviate significantly from most other values in a distribution. In this study, outlier detection is performed using the Z-Score method. The Z-Score measures how far a value is from the mean in terms of standard deviation units. A value is categorized as an outlier if it produces a z-score greater than +3 or less than -3 (McClave & Sincich, 2018).

2. Spatio-Temporal Median Polish Modeling

This study utilizes a two-way median polish model, which can be expressed as equation (1).

$$\mu(\boldsymbol{s},t) = \mu + \alpha_s + \tau_t + e \tag{1}$$

where: $\{\mu(\boldsymbol{s},t), s = 1,\dots,r;\ t = 1,\dots,c;\ \alpha_s \in \mathrm{D}|\mathrm{D} \subseteq \mathbb{R}^2, \tau_t \in \mathrm{T}\}, x \in R, r \geq 2$ is the number of rows, $c \geq 2$ is the number of columns, μ represents the general effects, α_l is the l-th row effects representing the spatial effects, dan τ_t is the t-th column effects representing the time effects, e is the error term in the median polish model. If the superscript in parentheses denotes the iteration cycle, $A_s^{(m)}$ and $B_t^{(m)}$ indicate the median estimates of the effects at various levels of spatial and time variables omitted in the iteration $v = (1,2,\dots,m)$, then the estimated effects of the variables can be determined as equation (2) (Martínez et al., 2017).

$$\begin{aligned} &\hat{\alpha}_s = \sum_{v=1}^{m} A_s^{(v)} - \underset{s}{med}\left\{\sum_{v=1}^{m} A_s^{(v)}\right\};\ \hat{\tau}_t = \sum_{v=1}^{m} B_t^{(v)} - \underset{t}{med}\left\{\sum_{v=1}^{m} B_t^{(v)}\right\}; \\ &\hat{\mu} = \underset{s}{med}\left\{\sum_{v=1}^{m} A_s^{(v)}\right\} + \underset{t}{med}\left\{\sum_{v=1}^{m} B_t^{(v)}\right\};\ s = 1,2,\dots,r; t = 1,2,\dots,c \end{aligned} \tag{2}$$

3. Kriging Modeling on Median Polish Spatial Effects

Kriging is a method used to predict the value of a random function, $\alpha(\boldsymbol{s})$, at one or more unobserved locations, based on n observed values within the domain D. The model can be expressed as equation (3) (Montero et al., 2015).

$$\hat{\alpha}(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i \alpha(\boldsymbol{s}_i) \tag{3}$$

where: $\alpha(\boldsymbol{s_i})$ are the observed median polish spatial effects values at n locations around s_0 (the prediction point) dan λ_i are the kriging weights, determined by assuming a zero mean and minimizing the variance of the prediction errors. There are three general types of kriging: simple kriging, ordinary kriging, and universal kriging.

a. Stationary in Spatial Data
   Spatial stationarity refers to the concept that the data depends only on the spatial characteristics and not on the specific locations $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$. The semivariogram is a tool used to evaluate the non-stationarity of spatial data. Spatial trends can be identified by plotting data against longitude and latitude for each time step (Rohma et al., 2023). Spatial non-stationarity can be addressed using the Box-Cox transformation, with data requiring adjustment if $\boldsymbol{\lambda} \neq \mathbf{1}$ or falls outside the 95% confidence interval.
b. Empirical Semivariogram
   The semivariogram estimate derived using the method of moments is expressed in equation (4) (Verdin et al., 2016).

$$\hat{\gamma}_\alpha(\boldsymbol{h}) = \frac{1}{2\#N(\boldsymbol{h})} \sum_{N(\boldsymbol{h})} \left(Z(\alpha_i + \boldsymbol{h}) - Z(\alpha_i)\right)^2 \tag{4}$$

   where $\#N(\boldsymbol{h})$ is the number of location pairs separated by the vector $\boldsymbol{h}$.
c. Theoretical Semivariogram
   Some common ones semivariogram for precipitation cases are exponential and gaussian models with functions according to equations (5) and (6) (Montero et al., 2015).
   1) Exponential model

$$\gamma_\alpha(|\boldsymbol{h}|) = c_0 + c\left\{1 - exp\left(-\frac{|\boldsymbol{h}|}{a}\right)\right\} \tag{5}$$

   2) Gaussian model

$$\gamma_\alpha(|\boldsymbol{h}|) = c_0 + c\left\{1 - exp\left(-\frac{|\boldsymbol{h}|^2}{a^2}\right)\right\} \tag{6}$$

   where: $\gamma_\alpha(|\boldsymbol{h}|)$ is semivariance at lag or distance h; c is a priori variability of the autocorrelation process; c_0 is nugget, representing spatially uncorrelated variance at distances smaller than the measurement error or sampling interval; a is *range*, the distance over which spatial autocorrelation or dependence persists. Value closer than a are correlated, while those beyond a are uncorrelated; and $c_0 + c$ is *sill*, representing the point where the variance stabilizes at a constant value.

d. Cross Validation
   A smaller forecast error indicates a better model for making predictions (Hu & Shu, 2019). One criterion used to evaluate forecast errors in this study is the Sum of Squared Errors.

4. ARIMA Modeling on Median Polish Time Effects

ARIMA (Autoregressive Integrated Moving Average) is a statistical algorithm used for forecasting univariate data. General form of ARIMA(p, d, q)(P, D, Q)^S can be seen in equation (7) (Cryer & Chan, 2008).

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B)^D\tau_t = \theta_0 + \theta_q(B)\Theta_Q(B^S)a_t \quad (7)$$

with,

$\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ is a non-seasonal AR polynomial orde p
$\Phi_P(B^s) = (1 - \Phi_1 B^S - \cdots - \Phi_P B^{Ps})$ is a seasonal AR polynomial orde P
$\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ is a non-seasonal MA polynomial orde q
$\Theta_Q(B^s) = (1 - \Theta_1 B - \cdots - \Theta_Q B^{Qs})$ is a seasonal MA polynomial orde Q
t=1,2,3,..., T, where T is the number of observations
(1-B)^d : non-seasonal differentiation with order d
(1-B)^D : seasonal differentiation with order D
a_t : residual at time-t

The steps that need to be taken in forming an ARIMA model consist of several stages starting from creating stationary data to forecasting (Ning et al., 2022).

a. Creating Stationary Data
   1) Stationary in variance
      Non-stationarity in variance can be overcome using the Box-Cox transformation. Data needs to be transformed if $\lambda \neq 1$ or if not at the 95% confidence interval.
   2) Stationary in mean
      The test that can be used to check the non-stationarity of data in mean is the unit root test (Augmented Dickey-Fuller Test) (Sarker & Khan, 2020). If the data is not stationary in mean, then it can be done differencing ($\Delta\tau_t = \tau_t - \tau_{t-1}$) .

b. Plotting ACF, PACF, and Model Identification
   ACF is a linear relationship between one variable and itself. PACF is used to see the correlation between variables at timet and variable in time $t-k$. ARIMA model identification is done by looking at significant ACF and PACF plot patterns to identify the order p, q, P, dan Q.
c. Parameter Estimation
   ARIMA parameter estimation can be done using maximul likelihood estimation methods.
d. Parameter significance
   The parameter significance test aims to test the feasibility of the model parameters.
e. Diagnostics residuals
   1) Residuals are normally distributed
      Residual normality testing can use Shapiro Wilk test.
   2) Non autocorrelation in residuals
      Non-autocorrelation testing between residuals can use the Ljung Box test.

f. Model Accuracy and Forcasting

The smaller the forecast error value produced by a model, the better it will be used to predict future periods. The forecasting error criteria that can be used are RMSE (Mean Square Error) and MAPE (Mean Absolute Percentage Error) (Islam et al., 2024). Model accuracy can be seen from the MAPE value which can be interpreted or interpreted into 4 categories, namely: <10% = very accurate, 10-20% = good, 20-50% = worthy, and >50% = inaccurate (Ruslana et al., 2024). Predictions for t the next stage correspond to the model produced with the equation (8) (Cryer & Chan, 2008).

$$\hat{\tau}_t = \phi_1\tau_{t-1} + \phi_2\tau_{t-2} + \cdots + \phi_p\tau_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{8}$$

5. Spatio Temporal Kriging Modeling on Median Polish Residuals

Spatio-temporal kriging is used to predict the unknown residual value $e(s_0, t_0)$ at an unobserved median polish residual spatial and time point (s_0, t_0). This prediction utilizes all available data related to regional variables, either from points across the entire domain or from smaller subsets, referred to as neighborhoods. A spatio-temporal random function is defined as $\{e(\boldsymbol{s}, t), \boldsymbol{s} \in D, t \in T\}$ where $D \subseteq \mathbb{R}^2$ and $T \subseteq \mathbb{R}$, and it is assumed that median polish residual have been observed at n spatial data $\times$ T time data $= N$ spatio-temporal data points $\{e(s_1, t_1), \dots, e(s_N, t_N)\}$. To estimate the value of a spatio-temporal median polish residual at the unobserved point (s_0, t_0), a linear predictor, as given in equation (9), is employed (Montero et al., 2015).

$$\hat{e}(\boldsymbol{s}_0, t_0) = \sum_{l=1}^{N} \lambda_l e(\boldsymbol{s}_l, t_l) \tag{9}$$

Here, $\hat{e}(\boldsymbol{s}_0, t_0)$ represents the predicted value at the unobserved point $(\boldsymbol{s}_0, t_0)$, λ_i are the spatio-temporal kriging weights, and $e(\boldsymbol{s}_l, t_l)$ are the observed values at known points. The weights in the spatio-temporal ordinary kriging model, which assumes an unknown mean and stationary data, can be determined using equation (10).

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \Sigma \end{pmatrix} = \begin{pmatrix} \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_1, t_1 - t_1) & \cdots & \gamma(\boldsymbol{s}_1 - \boldsymbol{s}_n, t_1 - t_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\boldsymbol{s}_N - \boldsymbol{s}_1, t_N - t_1) & \cdots & \gamma(\boldsymbol{s}_N - \boldsymbol{s}_N, t_N - t_N) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_e(\boldsymbol{s}_1 - \boldsymbol{s}_0, t_1 - t_0) \\ \vdots \\ \gamma_e(\boldsymbol{s}_N - \boldsymbol{s}_0, t_N - t_0) \\ 1 \end{pmatrix} \tag{10}$$

where N represents the number of spatial points in D, t is the time component, $\gamma_e(\boldsymbol{s}_i - \boldsymbol{s}_j, t_i - t_j)$ denotes the semivariogram between two observed spatio-temporal median polish residual, $\gamma_e(\boldsymbol{s}_j - \boldsymbol{s}_0, t_j - t_0)$ is the semivariogram between an observed median polish residual $(\boldsymbol{s}_i, t_i)$ and the unobserved median polish residual $(\boldsymbol{s}_0, t_0)$, and Σ is lagrange multiplier that associated with the condition of unbiasedness $(\sum_{i=1}^{N} \lambda_i = 1)$.

a. Stationary Data

Stationarity implies that the data depends only on spatial and temporal lags, without being influenced by specific spatials $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$ or the specific times $\boldsymbol{t_i}$ and $\boldsymbol{t_j}$ (Shand & Li, 2017).

1) Stationarity in spatial data
   The assessment of spatial stationarity for spatio-temporal data follows the same procedure as for purely spatial data (as described in Subsection 3, point a) and is conducted for each time step.
2) Stationarity in temporal data
   The temporal stationarity test is performed solely on the time component. To assess non-stationarity in panel data, the Im-Pesaran-Shin (IPS) unit root test can be utilized. The IPS statistic is calculated based on the average value of the Dickey-Fuller statistic for the n panel units (Murthy & Okunade, 2018). Im et al. (2003) provide the expected value and standard deviation for DF_{bar}. If the data is found to be non-stationary, differencing ($\Delta Z_t = Z_t - Z_{t-1}$) can be applied to achieve stationarity.

b. Empirical Spatio-Temporal Semivariogram
   The semivariogram estimate derived using the method of moments is given in equation (11) (de Medeiros et al., 2019).

$$\hat{\gamma}_e\big(\boldsymbol{h}(l), u(k)\big) = \frac{1}{2\#N\big(\boldsymbol{h}(l), u(k)\big)} \sum_{(\boldsymbol{s}_i, t_i), (\boldsymbol{s}_j, t_j) \in N\left(\boldsymbol{h}(l), u(k)\right)} \Big(e(\boldsymbol{s}_i, t_i) - e\big(\boldsymbol{s}_j, t_j\big)\Big)^2 \qquad (11)$$

   where, $N\big(\boldsymbol{h}(l), u(k)\big) = \big\{(\boldsymbol{s}_i, t_i)\big(\boldsymbol{s}_j, t_j\big) : \boldsymbol{s}_i - \boldsymbol{s}_j \in T\big(\boldsymbol{h}(l)\big), t_i - t_j \in T\big(u(k)\big)\big\}$, $\mathrm{T}\big(\boldsymbol{h}(l)\big)$ is the tolerance area in $\mathbb{R}^d$ around $\boldsymbol{h}(l)$ and $\mathrm{T}\big(u(k)\big)$ is the tolerance area in $\mathbb{R}$ around $u(k)$. $\#N\big(\boldsymbol{h}(l), u(k)\big)$ represents the number of distinct elements in the set $N\big(\boldsymbol{h}(l), u(k)\big)$, with $l = 1,2 \ldots, L$ and $k = 1,2, \ldots, K$.

c. Theoretical Spatio-Temporal Semivariogram
   Several semivariogram models used in this study are as follows (Montero et al., 2015).
   1) Product model or separable model
      The separable model is defined by equation (12).

$$\gamma_e(\boldsymbol{h}, u) = C_t(\boldsymbol{0})\gamma_s(\boldsymbol{h}) + C_s(\boldsymbol{0})\gamma_t(u) - \gamma_s(\boldsymbol{h})\gamma_t(u)), (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R} \qquad (12)$$

   2) Product-sum model
      The product-sum model is expressed as equation (13).

$$\gamma_e(\boldsymbol{h}, u) = \big(k_2 + k_1 C_t(\boldsymbol{0})\big)\gamma_s(\boldsymbol{h}) + \big(k_3 + k_1 C_s(\boldsymbol{0})\big)\gamma_t(u) - k_1\gamma_s(\boldsymbol{h})\gamma_t(u) \qquad (13)$$

      where, C_s and C_t is a covariance function, γ_s and γ_t is the semivariogram, $k_1, k_2, k_3 > 0$ and $k_1 + k_2 + k_3 > 0$ are constants. $C_s(\boldsymbol{0})$ and $C_t(\boldsymbol{0})$ are the sills from γ_s and γ_t, respectively.
   3) Metric model
      The metric model is defined as equation (14).

$$\gamma_e(\boldsymbol{h}, u) = \gamma\sqrt{\|\boldsymbol{h}\|^2 + (k.|u|)^2}, \ \ (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, c > 0 \qquad (14)$$

where $\|\boldsymbol{h}\| + c|u|$ represents the distance in $\mathbb{R}^d \times \mathbb{R}$ and c is a positive contant.

4) Sum-metric model

The semivariogram for the sum-metric model is given by equation (15).

$$\gamma_e(\boldsymbol{h}, u) = \gamma_s(\boldsymbol{h}) + \gamma_t(u) + \gamma\sqrt{\|\boldsymbol{h}\|^2 + (k.|u|)^2}, (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, c > 0 \quad (15)$$

5) Simple sum-metric model

The simple sum-metric model follows a separable form, as shown in equation (16).

$$\gamma_e(\boldsymbol{h}, u) = c_0 + \gamma_s(\boldsymbol{h}) + \gamma_t(u) + \gamma\sqrt{\|\boldsymbol{h}\|^2 + (k.|u|)^2}, (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}, c > 0 \quad (16)$$

d. Cross Validation

The optimal model is determined by the forecast error value, with smaller errors indicating better predictive performance. A common criterion for evaluating forecast error is the Root Mean Square Error (RMSE) (Islam et al., 2024).

6. Spatio Temporal Median Polish Kriging with ARIMA Integration Modeling

This model integrates spatial effects (via kriging) and time effects (via ARIMA), allowing for robust prediction and handling of spatio-temporal data by incorporating the spatio-temporal structure of the median polish model and spatio-temporal kriging, as described in equation (9), the spatial effects are added to the median polish model based on the kriging model (equation 3). Similarly, the time effects are incorporated into the median polish model based on the ARIMA model (equation 7). This results in the spatio-temporal median polish kriging model with ARIMA integration, expressed as equation (17).

$$\hat{Z}(\boldsymbol{s_0}, t_0) = \hat{\mu} + \sum_{i=1}^{n} \lambda_i \alpha(\boldsymbol{s}_i) + \hat{\tau}_t + \sum_{i=1}^{n}\sum_{t=1}^{T} \lambda_{it} e(s_i, t_t) \quad (17)$$

with $\hat{\tau}_t = \frac{\widehat{\theta}_0 + \widehat{\theta}_q(B)\widehat{\Theta}_Q(B^S)a_t}{\widehat{\Phi}_P(B^s)\widehat{\phi}_p(B)(1-B)^d(1-B)^D}$, $e(s_i, t_i)$ is the residuals from the median polish model. The flow chart presented in Figure 1 below shows the flow of the research method.

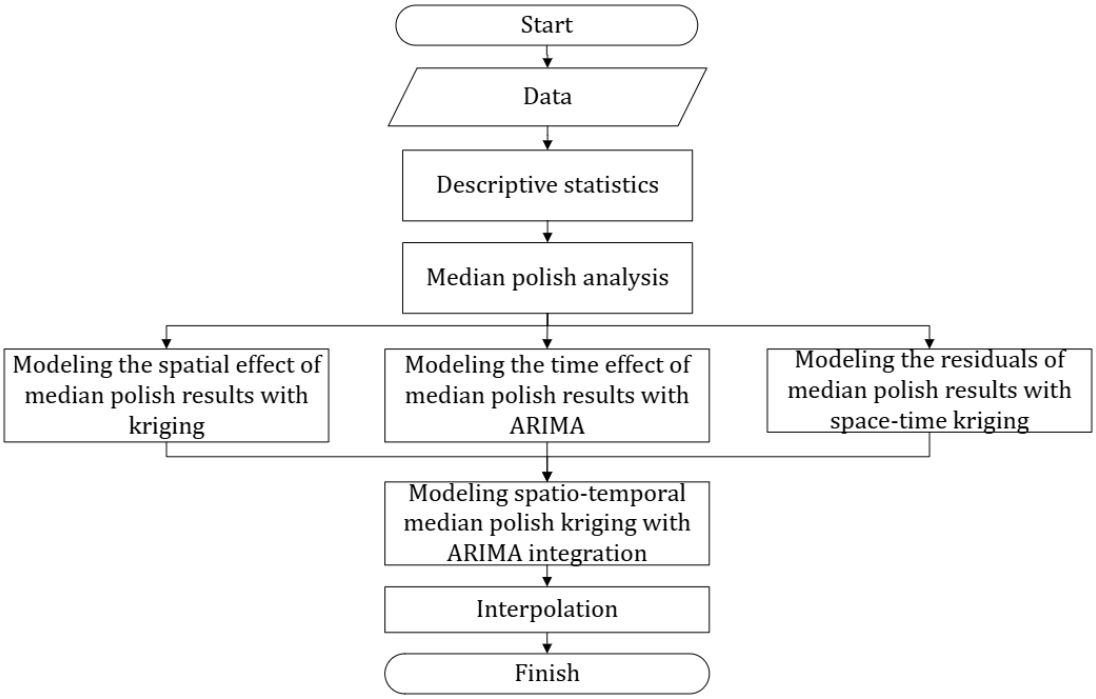


Figure 1. Flow Chart

C. RESULT AND DISCUSSION

1. Outlier Checking

Using the Z-Score method, precipitation in West Kutai in May 2022 was identified as an outlier, with a Z-Score of 4.247 exceeding the outlier threshold of 3. Outliers in precipitation data can significantly impact interpolation accuracy, especially in models that assume normality, such as ordinary kriging. The presence of extreme values may lead to biased estimates, reducing the reliability of spatial and temporal predictions. To address this issue, median polish kriging was adopted due to its robustness against outliers. By decomposing the data into additive components and applying a median-based approach, this method minimizes the influence of extreme values while preserving the underlying spatio-temporal patterns of precipitation. This makes it particularly suitable for datasets with high variability, such as precipitation observations

2. Spatio-Temporal Median Polish

Median polish was applied by iteratively subtracting the median of each row and column until the median value reached zero. The results yielded a general mean (μ) of 230.775 mm, with spatial effects, time effects, and residuals presented in Table 1, 2, and 3, respectively.

Table 1. Median Polish Spatial Effects (α_s)

Location	Temindung Station	Sultan Aji Station	Kalima-rau Station	Sangkuli-rang Post	Kembang Jenggut Post	Long Iram Post	Gunung Telihan Post
Precipitation Effect (mm)	-12.400	-1.700	-32.375	-65.525	8.725	69.175	-0.85

Table 2. Median Polish Time Effects (τ_t)

Time	1*	2	3	4	5	6
Precipitation Effect (mm)	-184	-279.575	-220.725	-265.725	-247.875	-321.725
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Time	31	32	33	34	35	36**
Precipitation Effect (mm)	-311.875	-375.275	-314.850	-320.275	-254.925	-230.775

*January 2021; **December 2023

Table 3. Median Polis Residuals (e)

Residuals (mm)		**Time**					
		1	**2**	**3**	**4**	⋯	**36**
Location	Temindung Station	103.850	-21.575	45.575	72.575	⋯	-59.675
	Sultan Aji Station	-25.450	-44.875	-72.125	-34.325	⋯	37.625
	⋮	⋮	⋮	⋮	⋮	⋱	⋮
	Gunung Telihan Post	0	12.575	123.225	-0.675	⋯	50.075

After applying median polish, an outlier was identified in the residuals. Specifically, the residuals of precipitation in West Kutai during May 2022 yielded a Z-score of 3.581. This indicates that median polish does not entirely eliminate extreme outliers, but can mitigate their impact on the model.

3. Kriging Modeling on Median Polish Spatial Effects

Kriging was applied to the median polish spatial effects data presented in Table 1. The results, illustrated in Figure 2, indicate a trend in the precipitation spatial effects based on longitude (Easting), where the spatial effects decrease progressively further eastward. Given this coordinate-based trend, universal kriging was selected as the appropriate method. These findings demonstrate that median polish can addresses trends or non-stationary data in spatio-temporal datasets by decomposing their components.

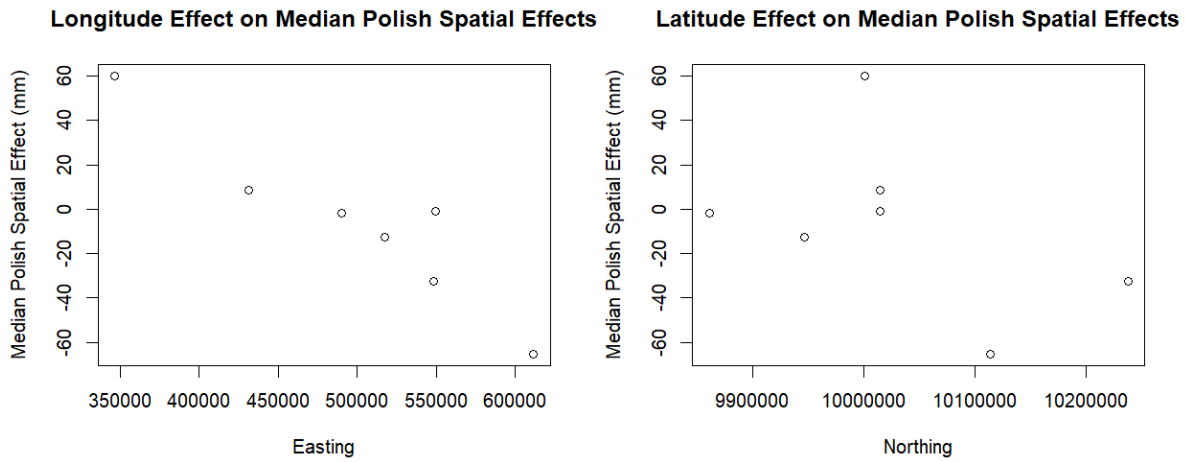


Figure 2. Median Polish Spatial Effects Based on Longitude and Latitude

The kriging results are highly dependent on the selected semivariogram model. Table 4 shows the parameter estimates and SSE generated from the three models used.

Table 4. Fit Semivariogram Model for Location Effects Median Polish

Model	Nugget	Partial sill	Effective Range	SSE
Spherical	164.844	251.272	46,328.610	0.000056
Eksponential	284.452	214.431	166,796.600	0.000048
Gaussian	265.098	151.734	68,613.150	0.000051

Based on the SSE value shown in Table 4, exponential model produced the lowest sum of squares error. Using this model, the following parameters were obtained:

a. Nugget (284.452): (1) a relatively high nugget value in this study (284.452 m) indicates significant precipitation variability at small scales or noise in the data that cannot be captured by the exponential semivariogram model; and (2) high nugget value may cause the interpolation results to be smoother as the model balances between small-scale variability and larger spatial correlation.
b. Sill (499): (1) a high sill value indicates considerable differences in precipitation between locations, but there is still a spatial pattern that can be analyzed before reaching the maximum range; and (2) high sill value indicates significant precipitation variability in the study area, leading to higher variation in interpolation predictions between locations.
c. Effective Range and Range (166,796.6 and 55,598.867): (1) in the exponential model, the effective range is calculated as $3a$, where a represents the range. This implies that if two locations are within 55,59 km of each other, their precipitation measurements are still strongly correlated. However, beyond this distance, spatial correlation is no longer significant, making kriging predictions less accurate for locations far from observation points; and (2) a large range suggests that precipitation has a broad spatial pattern, allowing interpolation to be performed while considering spatial relationships over long distances.

Based on the parametre estimation that has been obtained, a semivariogram model can be formed as follows.

$$\gamma_{\alpha}(|\boldsymbol{h}|) = 499\left\{1 - exp\left(-\frac{|\boldsymbol{h}|}{55{,}589.867}\right)\right\} \tag{18}$$

The theoretical exponential semivariogram is depicted in Figure 3, while the kriging results are visualized in Figure 3.

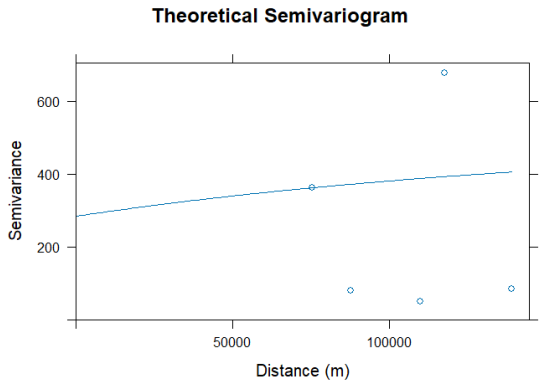


Figure 3. Theoretical Semivariogram

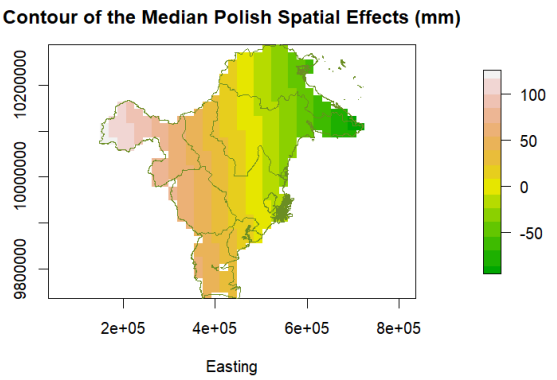


Figure 4. Contour of The Spatial Effects

The contour plot in Figure 4 highlights that precipitation increases westward, likely due to the region's distance from the ocean. Consistent with the water vapor theory, clouds carrying water vapor are pushed toward areas of higher altitude and lower air pressure, leading to greater precipitation in the western, more mountainous regions.

4. ARIMA Modelling on Median Polish Time Effects

Based on the median polish time effects plot in Figure 5, the median polish time effects fluctuate from January 2021 to December 2023, showing a decreasing trend from early to mid-year and an increase towards the end of the year. However, this pattern does not consistently repeat each year, indicating the absence of a stable seasonal pattern. Further analysis using 6-month and 12-month intervals presented in Figures 6 and Figure 7, also reveals no periodically recurring pattern. Median polish time effects fluctuations between periods do not exhibit the regularity characteristic of seasonal patterns. Although the 12-month interval shows a general trend of increase and decrease, the variations remain irregular. An examination with the Auto Correlation Function (ACF) in Figure 8 also does not indicate a strong seasonal pattern. There are no significant correlation spikes at repeated lags, which should be expected if an annual seasonal pattern were present. Most lags show a decreasing correlation without a recurring pattern, confirming that median polish time effects variations are influenced more by variables other than seasonality.

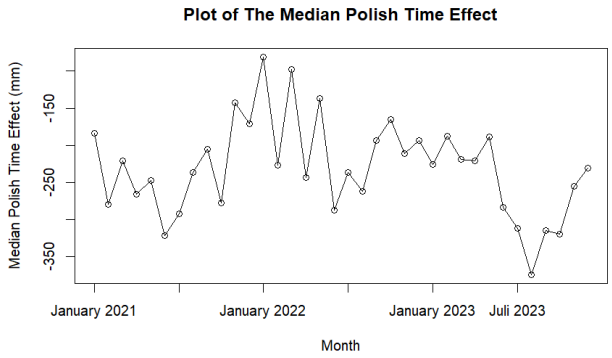


Figure 5. Plot of the Median Polish Time Effects

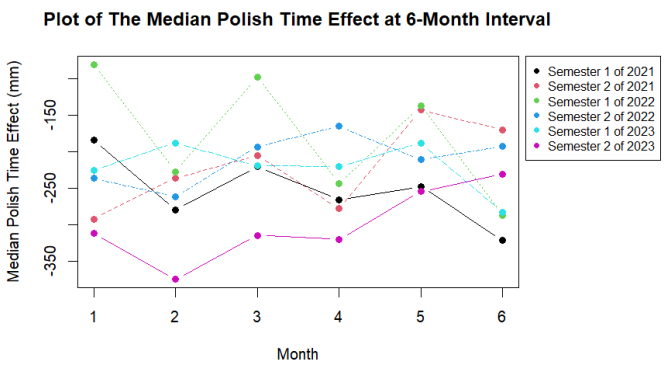


Figure 6. Plot of Median Polish Time Effects at 6-Months Interval

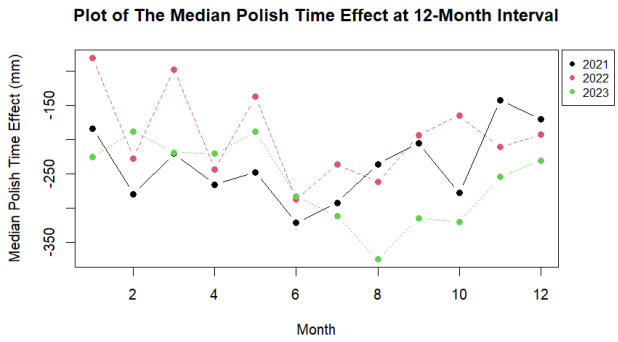


Figure 7. Plot of Median Polish Time Effects at 12-Months Interval

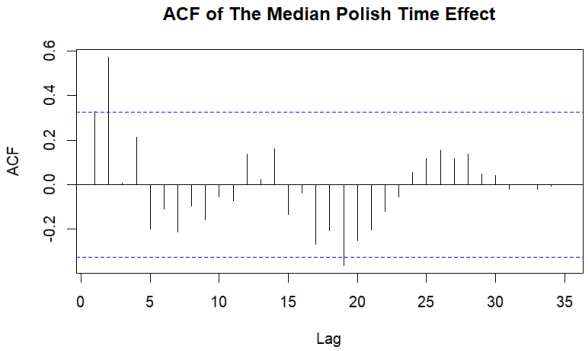
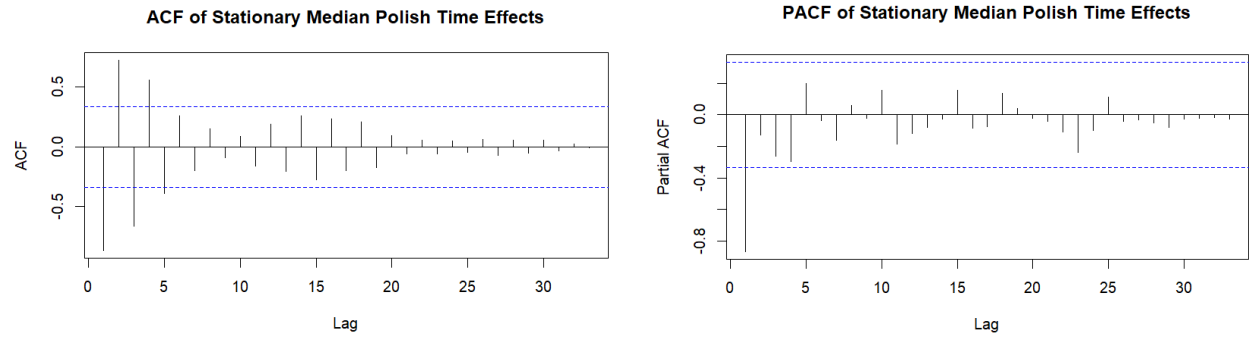


Figure 8. ACF of Median Polish Time Effects

The time effects obtained from the median polish results in Table 2 were forecasted using ARIMA. A prerequisite for ARIMA modeling is that the data must be stationary. The Box-Cox transformation confirmed stationarity in variance, but the Augmented Dickey-Fuller (ADF) test revealed non-stationarity in the mean. After applying second-order differencing, the data became stationary. The ACF and PACF plots in Figure 9 show that the PACF is significant only at lag 1, and the ACF exhibits a sinusoidal pattern without seasonality. Consequently, ARIMA(1,2,0) was chosen as the most suitable model. The estimated AR parameter ϕ was -0.881, allowing the ARIMA model for the time effects of median polish to be written as follows.

$$\tau_t = 2\tau_{t-1} - \tau_{t-2} - 0.881(\tau_{t-1} - 2\tau_{t-2} + \tau_{t-3}) + a_t \tag{19}$$



(a) (b)

Figure 9. (a) ACF (b) PACF of Stationary Time Effects

The AR parameter was statistically significant ($p < 0.05$), residuals followed a normal distribution, and no autocorrelation was present, confirming the model's appropriateness for forecasting time effects derived from the median polish process. With a MAPE of 31%, the model demonstrated good predictive performance. The forecasts are visualized in Figure 10.

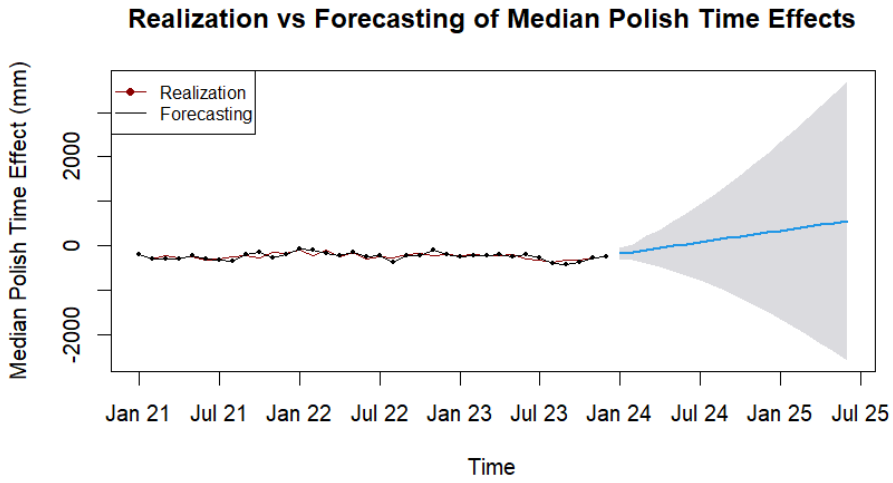


Figure 10. Realization vs Forcasting of Median Polish Time Effects

5. Spatio Temporal Kriging Modelling on Median Polish Model Residuals

Spatio-temporal kriging was applied to the stationary residuals from the median polish model. Spatial stationarity was confirmed through longitude and latitude plots and regression tests, revealing no significant trends across all observation times. This finding supports the premise that median polish can addresses non-stationary data. Among the various combinations of spatial, temporal, and joint semivariogram models, the simple sum-metric model with spatial, temporal, and joint semivariogram follow gaussian model emerged as the best-fitting semivariogram, producing the lowest RMSE of 2668.869. The parameters derived from this model are as follows:

a. Spatial component: (1) partial sill: 760,222.3, indicating the proportion of the total variance (sill) attributed to spatial autocorrelation. This reflects the degree to which nearby locations exhibit similar precipitation values; and (2) effective range: 4,800,758 m (approximately 4801 km), signifying the maximum distance over which spatial correlation exists between two locations.
b. Temporal Component: (1) partial sill: 0, indicating no autocorrelated spatial variability within the data over specific temporal distances; (2) effective range: 1,128,942 m (approximately 1129 km), signifying the maximum temporal distance over which spatial correlation exists.
c. Joint Component: (1) partial sill: 0, indicating no significant spatial or temporal variability captured jointly; and (2) effective range: 96,507 m (approximately 96.5 km), with a similar interpretation to the spatial and temporal components.

Additional parameters include:

a. Nugget: 6993.967 m, representing unexplained variability at a distance of zero, such as measurement error or micro-scale variability.
b. Anisotropy Correction: 3229.449 m, which adjusts for the disparity between spatial and temporal correlation scales. This suggests that one temporal unit is equivalent to a spatial distance of approximately 3.229 km.

In the Gaussian model, there is a relationship between the effective range and the parameter a (range) expressed as: effective range $= a\sqrt{3}$. Using the derived parameters, the semivariogram model can be formulated as follows:

$$\gamma_e(\boldsymbol{h}, u) = 6993.967 + \gamma_s(\boldsymbol{h}) + \gamma_t(\boldsymbol{u}) + \gamma_{st}\left(\sqrt{\|\boldsymbol{h}\|^2 + (3229.449|u|)^2}\right) \tag{20}$$

with,

$$\gamma_s(|\boldsymbol{h}|) = 760{,}222.3\left\{1 - exp\left(-\frac{|\boldsymbol{h}|^2}{2{,}771{,}718.9^2}\right)\right\}$$

$$\gamma_t(\boldsymbol{u}) = 0\left\{1 - exp\left(-\frac{|u|^2}{651{,}794.97^2}\right)\right\}$$

$$\gamma_{st}\left(\sqrt{\|\boldsymbol{h}\|^2 + (3229.449|u|)^2}\right) = 0\left\{1 - exp\left(-\frac{\|\boldsymbol{h}\|^2 + (3229.449|u|)^2}{55{,}718.34^2}\right)\right\}$$

The empirical and theoretical semivariograms for this model are depicted in Figure 11.

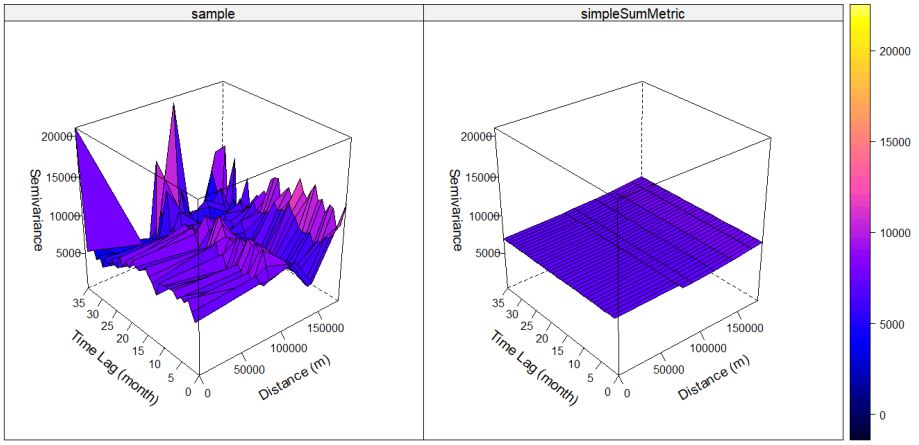


Figure 11. Sample and Simple Sum-Metric Semivariogram

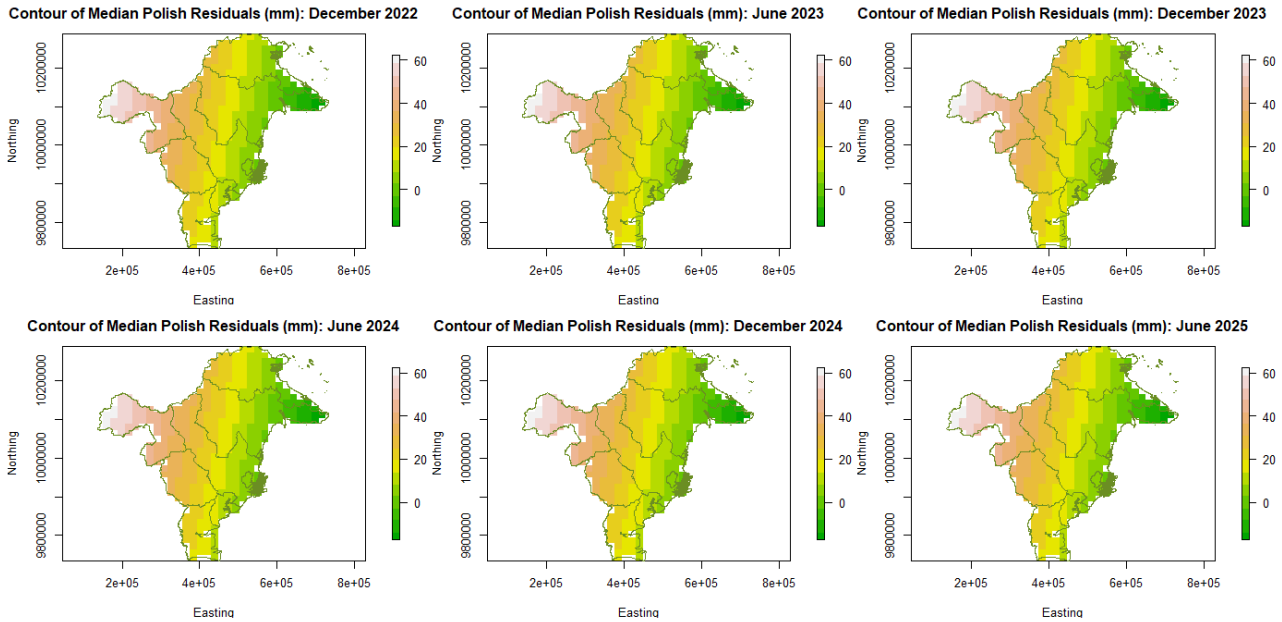


Figure 12. Contour of Median Polish Residuals

Interpolation using the best semivariogram produced the contours in Figure 12, which indicate that precipitation distribution patterns remain consistent annually. Precipitation increases westward, aligning with the spatial effects trend.

6. Spatio Temporal Median Polish Kriging with ARIMA Integration Modeling

Spatio-temporal median polish kriging with ARIMA integration combined kriging results for general mean of 230.775, median polish spatial effects modeled using the semivariogram in equation (18), ARIMA forecasts for median polish time effects presented in equation (19), and spatio-temporal kriging for median polish residuals using the semivariogram model in equation (20). The final model is expressed in equation (21) and produces the interpolated precipitation for East Kalimantan, as shown in Figure 13.

$$\hat{Z}(\boldsymbol{s_0}, t_0) = 230.775 + \sum_{i=1}^{7} \lambda_i Z(\boldsymbol{s}_i) + 2\tau_{t-1} - \tau_{t-2} - 0.881(\tau_{t-1} - 2\tau_{t-2} + \tau_{t-3}) + a_t + \sum_{l=1}^{252} \lambda_l e(s_l, t_l) \tag{21}$$

The final interpolated results (Figure 13) reveal that precipitation increases spatially toward the west, while latitude shows no significant effect, resulting in relatively uniform precipitation along the same latitude. Temporally, precipitation demonstrates an upward trend, with fewer regions experiencing low to medium precipitation (0–300 mm).

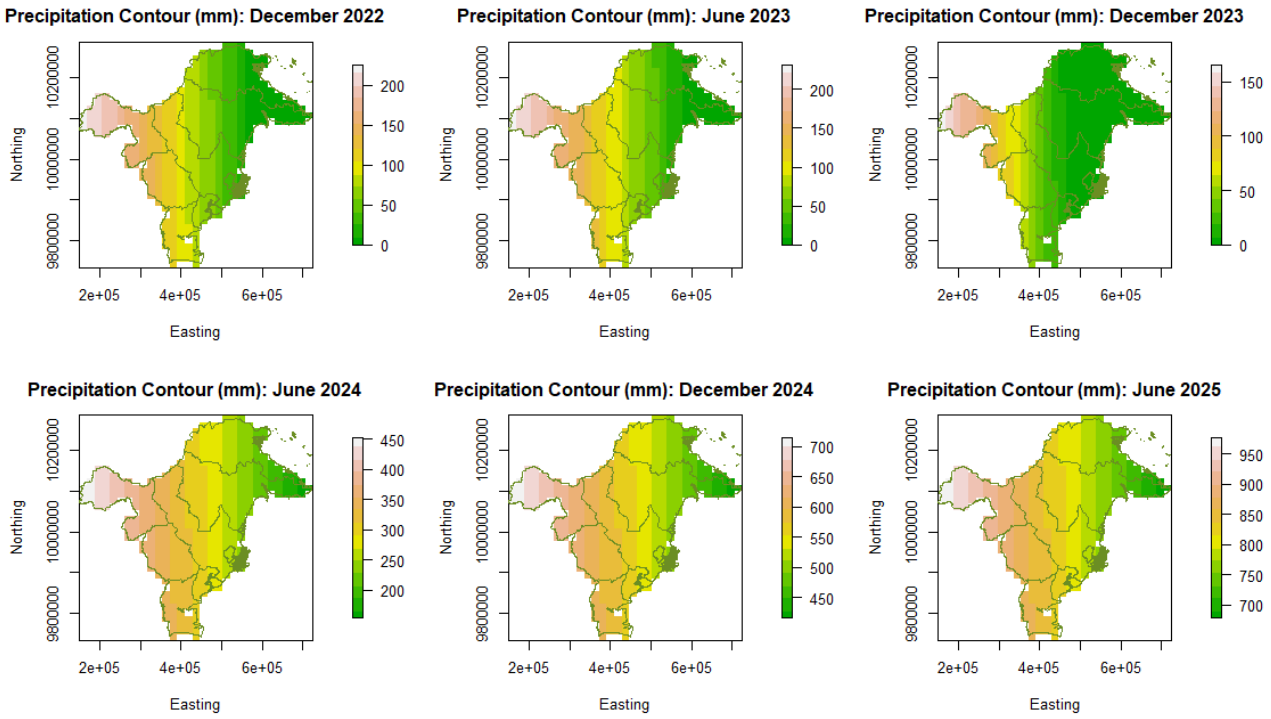


Figure 13. Contour of Precipitation

D. CONCLUSION AND SUGGESTIONS

This study concludes that median polish can reduces the influence of outliers. The optimal kriging model identified for spatio-temporal median polish kriging consists of an exponential model for spatial kriging and a simple sum-metric model for spatio-temporal kriging, with spatial, temporal, and joint semivariograms following the gaussian model. Additionally, the best ARIMA model for median polish time effects in this study is ARIMA(1,2,0). The results indicate that western East Kalimantan experiences higher precipitation compared to the eastern region. By December 2024, precipitation levels in East Kalimantan are projected to increase, necessitating proactive mitigation measures to reduce the risks of flooding and landslides. Furthermore, effective water conservation strategies must be implemented to minimize the potential adverse impacts of increased precipitation. However, this study has certain limitations. The use of only seven observation posts may lead to increased uncertainty in interpolated values, especially in areas far from measurement stations. This could reduce the reliability of precipitation estimates in regions with high variability. Additionally, the chosen semivariogram model may not fully capture complex spatial and temporal dependencies, potentially introducing bias in kriging predictions. Alternative semivariogram models, such as Matérn or Cauchy models, could be explored to better accommodate long-range spatial correlations and anisotropic patterns. The use of ARIMA for modeling the time effects may lead to less accurate predictions, as it does not account for the influence of other variables affecting the median polish time effects. Furthermore, incorporating additional climate variables, such as wind speed, humidity, and sea surface temperature, could enhance model performance by accounting for key meteorological drivers of precipitation variability. These limitations suggest careful interpretation of results, particularly in data-sparse areas. Future research should expand the dataset, explore alternative semivariogram models, integrate additional climate variables to improve model accuracy by using time series models that consider predictor variables, such as the ARIMAX model.

REFERENCES

Adiguna, I. B. A. P., Nuarsa, I. W., & Puspitha, N. L. P. R. (2021). Pengaruh Suhu Permukaan Laut terhadap Curah Hujan di Pulau Bali Tahun 2009-2018. *Journal of Marine and Aquatic Sciences*, *7*(2), 214. https://doi.org/10.24843/jmas.2021.v07.i02.p10

Barbara, D., & Wu, X. (2003). An Approximate Median Polish Algorithm for Large Multidimensional Data Sets. *Knowledge and Information Systems*, *5*(4), 416–438. https://doi.org/10.1007/s10115-003-0096-7

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis with Applications in R* (2nd ed.). Springer. https://doi.org/https://doi.org/10.1007/978-0-387-75959-3

de Medeiros, E. S., de Lima, R. R., de Olinda, R. A., Dantas, L. G., & dos Santos, C. A. C. (2019). Space-time kriging of precipitation: Modeling the large-scale variation with model GAMLSS. *Water (Switzerland)*, *11*(11), 1–16. https://doi.org/10.3390/w11112368

Han, L., Wang, L., Chen, H., Xu, Y., Sun, F., Reed, K., Deng, X., & Li, W. (2022). Impacts of Long-Term Urbanization on Summer Rainfall Climatology in Yangtze River Delta Agglomeration of China. *Geophysical Research Letters*, *49*(13), 1–12. https://doi.org/10.1029/2021GL097546

Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About Regression-Kriging: From Equations to Case Studies. *Computers and Geosciences*, *33*(10), 1301–1315. https://doi.org/10.1016/j.cageo.2007.05.001

Hu, D. G., & Shu, H. (2019). Spatiotemporal interpolation of precipitation across Xinjiang, China using space-time CoKriging. *Journal of Central South University*, *26*(3), 684–694. https://doi.org/10.1007/s11771-019-4039-1

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*(1), 53–74. https://doi.org/10.1016/S0304-4076(03)00092-7

Islam, N., Kashem, A., Das, P., Ali, M. N., & Paul, S. (2024). Prediction of high-performance concrete compressive strength using deep learning techniques. *Asian Journal of Civil Engineering*, *25*(1), 327–341. https://doi.org/10.1007/s42107-023-00778-z

Katipoğlu, O. M. (2022). Spatial analysis of seasonal precipitation using various interpolation methods in the Euphrates basin, Turkey. *Acta Geophysica*, *70*(2), 859–878. https://doi.org/10.1007/s11600-022-00756-0

Martínez, W. A., Melo, C. E., & Melo, O. O. (2017). Median Polish Kriging for space–time analysis of precipitation. *Spatial Statistics*, *19*(1), 1–20. https://doi.org/10.1016/j.spasta.2016.10.003

McClave, J., & Sincich, T. (2018). *Statistics*. Pearson.

Montero, J.-M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging* (1st ed.). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118762387

Murthy, V., & Okunade, A. (2018). Is the health care price inflation in US urban areas stationary?: Evidence from panel unit root tests. *Journal of Economics, Finance and Administrative Science*, *23*(44), 77–94. https://doi.org/10.1108/JEFAS-02-2017-0043

Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers and Geosciences*, *164*(1), 1–11. https://doi.org/10.1016/j.cageo.2022.105126

Pertiwi, M. A., Kahar, S., Sasmito, B., & Marpaung, S. (2015). Analisis Korelasi Suhu Permukaan Laut Terhadap Curah Hujan dengan Metode Penginderaan Jauh Tahun 2012-2013 (Studi Kasus : Kota Semarang). *Jurnal Geodesi Undip Januari*, *4*(1), 61–71. https://doi.org/https://doi.org/10.14710/jgundip.2015.7639

Rohma, N. N., Pramoedyo, H., & Astutik, S. (2023). Perbandingan Pendugaan Metode Ordinary Kriging dan Metode Ordinary Kriging dengan Teknik Jackknife. *MAp (Mathematics and Applications) Journal*, *2*(4), 101–111. https://doi.org/https://doi.org/10.15548/map.v4i2.4736

Ruslana, Z. N., Prihatin, R. S., Sulistiyowati, S., & Nugroho, K. (2024). Application of the Arima Method to Prediction Maximum Rainfall at Central Java Climatological Station. *Sinkron*, *8*(4), 2135–2141. https://doi.org/10.33395/sinkron.v8i4.13984

Sari, F. P., & Atsidiqi, S. N. (2020). Investigasi Pengaruh Skenario Modifikasi Urbanisasi pada Perubahan Hujan di Kota Makassar. *Prosiding SNFA (Seminar Nasional Fisika Dan Aplikasinya)*, *5*(1), 85–96. https://doi.org/10.20961/prosidingsnfa.v5i0.46597

Sarker, B., & Khan, F. (2020). Nexus between foreign direct investment and economic growth in Bangladesh: an augmented autoregressive distributed lag bounds testing approach. *Financial Innovation*, *6*(1), 1–18. https://doi.org/10.1186/s40854-019-0164-y

Shand, L., & Li, B. (2017). Modeling nonstationarity in space and time. *Biometrics*, *73*(3), 759–768. https://doi.org/10.1111/biom.12656

Sudinda, T. W. (2020). Pemanfaatan Air Hujan untuk Memenuhi Kebutuhan Air Baku Jangka Panjang Ibu Kota Negara. *Construction Engineering and Sustainable Development*, *3*(1), 1–48. https://www.e-journal.trisakti.ac.id/index.php/sipil/article/view/13654

Sun, Y., & Genton, M. G. (2012). Functional Median Polish. *Journal of Agricultural, Biological, and Environmental Statistics*, *17*(3), 354–376. https://doi.org/10.1007/s13253-012-0096-8

Tutmez, B. (2014). Analyzing non-stationarity in cement stone pit by median polish interpolation: a case study. *Journal of Applied Statistics*, *41*(2), 454–466. https://doi.org/10.1080/02664763.2013.840274

Verdin, A., Funk, C., Rajagopalan, B., & Kleiber, W. (2016). Kriging and local polynomial methods for blending satellite-derived and gauge precipitation estimates to support hydrologic early warning systems. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(5), 2552–2562. https://doi.org/10.1109/TGRS.2015.2502956

Yang, X., Yang, Y., Li, K., & Wu, R. (2020). Estimation and characterization of annual precipitation based on spatiotemporal kriging in the Huanghuaihai basin of China during 1956–2016. *Stochastic Environmental Research and Risk Assessment*, *34*(9), 1407–1420. https://doi.org/10.1007/s00477-019-01757-0

Zhang, W., Liu, D., Zheng, S., Liu, S., Loáiciga, H. A., & Li, W. (2020). Regional precipitation model based on geographically and temporally weighted regression kriging. *Remote Sensing*, *12*(16), 1–19. https://doi.org/10.3390/RS12162547