Science Essence Journal Vol. 42 No. 1 (2026) | 131-152

Research Article

Quantitative Assessment of Interpolation Methods for Handling Missing Data in Long-Term Tropical Meteorological Time Series

Novi Reandy Sasmita[1*], Novita Sari Saragih[1], Latifah Rahayu[1], Feby Apriliansyah[2], Arinda Ma-a-lee[3] and Muhammad Farid[1]

Received: 18 January 2026
Revised: 22 March 2026
Accepted: 2 April 2026

ABSTRACT

The integrity of long-term tropical meteorological time series is crucial for climate modeling, yet missing data compromises dataset reliability. This study aimed to identify the most accurate interpolation method for handling missing values in key meteorological variables to support robust climate research and action. This study conducted a simulation study using a 14-year daily time series (2010–2023) from North Aceh, Indonesia, comprising 25,565 observations across five variables: temperature, humidity, rainfall, sunshine duration, and wind speed. The baseline dataset was subjected to simple random missingness at 10%, 20%, and 30%. Four interpolation methods: linear, spline, stineman, and moving average were applied. The performance of these methods was rigorously evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Scaled Error (MASE). Linear interpolation consistently demonstrated superior performance, yielding the lowest error rates across all variables and missing data percentages. In contrast, spline and moving average methods exhibited higher error metrics and greater sensitivity to outliers, particularly for volatile variables like rainfall. Stineman interpolation showed moderate, intermediate performance. For long-term tropical meteorological data, linear interpolation is the most effective and stable method for imputing missing values. This finding provides a validated, practical solution for enhancing the quality of climate records, a fundamental requirement for achieving the monitoring and analysis goals of Sustainable Development Goal 13 on Climate Action. The study recommends adopting this method for creating reliable datasets for climate trend analysis and modeling.

Keywords: Interpolation, Meteorological data, Missing data, Long-term tropical data

[1] Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[2] Department of Regional and Urban Planning, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
[3] Department of Mathematics and Computer Science, the Faculty of Science and Technology, Prince of Songkla University, Pattani 94000, Thailand
*Corresponding author, email: novireandys@usk.ac.id

Introduction

Interpolation is an essential analytical method used to estimate missing data points within the range of existing values. This method is especially critical in meteorology, where the completeness of datasets significantly impacts the accuracy of models and predictions [1]. Reliable meteorological forecasts are heavily dependent on uninterrupted data streams; thus, addressing data gaps is imperative for ensuring robust analyses. Missing meteorological data often results from factors such as equipment malfunctions, sensor breakdowns, or errors during data recording [2].

This challenge is amplified in tropical regions. The region's pronounced climatic heterogeneity, marked by high variability and sharply defined wet and dry periods, is often complicated by rapid, extreme meteorological shifts. This environmental volatility poses a significant challenge to data integrity, increasing the incidence of data voids and simultaneously complicating their accurate interpolation. The underlying non-uniformity of the data structure makes conventional gap-filling methods less reliable. Consequently, the insufficient management of these missing values risks obscuring or misrepresenting the intrinsic dynamics of the local climate system, thereby impeding the development of robust climatic models and effective data-driven mitigation strategies [3]. Variables like temperature, humidity, rainfall, sunshine duration, and wind speed are particularly susceptible to these disruptions, often leading to reduced accuracy in predictions critical to sectors like agriculture, transportation, and tourism. Implementing appropriate interpolation methods helps retain the integrity and usability of meteorological datasets, mitigating potential biases that incomplete data can introduce [4].

Over the years, various interpolation methods have been developed to address missing meteorological data, each with distinct strengths and limitations. Previous studies indicate that interpolation accuracy depends more on the statistical characteristics of the variable than on algorithmic complexity alone. Linear interpolation is simple and computationally efficient, and it often performs well for relatively stable series [5, 6]. Spline interpolation is useful for smooth curves but may overshoot in highly variable data [7–9]. Stineman interpolation preserves local shape and reduces oscillation [10, 11], whereas moving average interpolation smooths short-term noise but may blur abrupt changes [12, 13]. Therefore, this study compares these methods under controlled missingness to identify the most reliable approach for long-term tropical meteorological time series.

The application of linear interpolation in meteorology allows for the estimation of missing values based on the known values at adjacent time points or spatial locations. This method is particularly advantageous due to its simplicity and computational efficiency, making it suitable for real-time data processing and analysis. Previous discusses the importance of integrating spatial and temporal data to improve the precision of meteorological datasets, suggesting that linear interpolation can be part of a broader spatiotemporal interpolation framework. Further, linear and spline functions are used to interpolate the weather data for filling missing meteorological data in heating and cooling seasons separately of Kerman, Iran. Both methods provide accurate results [14].

Spline interpolation is widely recognized for its ability to create smooth curves that align with the general trends of unevenly distributed data points. This method has demonstrated effectiveness in

filling gaps in temperature and precipitation datasets, ensuring continuity in data representation. For instance, spline interpolation has been particularly successful in rainfall modeling, even with incomplete datasets [9]. Another prominent approach, stineman interpolation, is designed to preserve the natural shape and trends of environmental data, making it well-suited for variables that exhibit specific patterns [10]. Although its application in meteorology is relatively limited, existing studies highlight its ability to reconstruct temperature variations accurately while minimizing distortions. Additionally, moving average interpolation offers a simpler approach, effectively smoothing data to reduce short-term fluctuations and emphasize longer-term trends [12]. This method, though less complex, remains a popular choice for preliminary data analyses in meteorological studies [13].

The primary challenge addressed in this study is the selection of the optimal interpolation method to enhance the accuracy of meteorological datasets with missing values. A comparative analysis of four interpolation methods, linear, spline, stineman, and moving average interpolation, is proposed as the general solution. This approach evaluates the relative performance of these methods across key meteorological variables to determine their effectiveness in filling data gaps. This study adopts a systematic methodology to explore the effectiveness of interpolation methods under varying levels of data incompleteness. Existing studies, while insightful, often focus on short-term datasets, leaving a gap in understanding their applicability to long-term meteorological data. This study bridges that gap by analyzing over a decade of meteorological records, offering a robust framework for future data analyses.

The specific solution involves a rigorous performance evaluation of the four interpolation methods using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Scaled Error (MASE). By simulating missing data scenarios at varying percentages, this study provides a comprehensive assessment of each method's reliability and accuracy. An extensive review of relevant literature reveals a study gap in the application of interpolation methods to long-term meteorological datasets. While prior studies have evaluated these methods in controlled environments or over short periods, their effectiveness in long-term analyses still needs to be explored. This study addresses this limitation by applying these methods to an extended dataset from North Aceh Regency, Indonesia.

The study aims to identify the optimal interpolation method for managing long-term missing meteorological data to enhance the quality and reliability of extended data sets. This study lies in the comprehensive comparison of several methods across different meteorological variables over 13 years. The findings are expected to advance current methodologies and provide practical recommendations to improve meteorological data analysis, which in turn can support informed decision-making in climate change-sensitive sectors. This study supports SDG 13 (Climate Action) by improving the completeness and reliability of long-term meteorological records, which are fundamental for climate monitoring, extreme-event analysis, and adaptation planning. By identifying the most reliable interpolation method for daily tropical meteorological time series, this study strengthens the quality of data used in climate trend assessment, agricultural planning, and disaster risk reduction. Thus, the contribution to SDG 13 is

both methodological and practical, as more complete and reliable climate archives strengthen the evidence base for climate action.

Materials and Methods

The Study Area, Data Source, and Variables

North Aceh, a prominent regency in Aceh Province as one of the tropical regions in Indonesia, holds the distinction of being the province's largest rice-producing region. This reason makes precise meteorological data crucial for supporting agricultural productivity in the area. In 2022, the population of North Aceh was recorded at 614,640, with a population density of 186.43 individuals per square kilometer. A significant portion of the population is engaged in agriculture. The selection of North Aceh Regency for this study was conducted through purposive sampling, considering its role as the largest contributor to rice production in Aceh Province, with a total yield of 238,088 tonnes in 2023 [15]. The focus on North Aceh, a key agricultural region. By improving the accuracy of meteorological data crucial for agriculture a sector highly vulnerable to climate variability, this work provides a tool for safeguarding food security, which is an integral component of climate adaptation and resilience-building efforts. The location of the study area is shown in Figure 1.

Figure 1 Location of North Aceh Regency showing the study sites.

The dataset for this study was obtained from the official Indonesian Agency for Meteorological, Climatological and Geophysics (BMKG) website (www.dataonline.bmkg.go.id) and comprises daily meteorological observations for North Aceh Regency, spanning from January 1, 2010, to December 31, 2023. The analyzed variables include temperature (°C), humidity (%), rainfall (mm), sunshine duration (hours), and wind speed (m/s). Each variable consists of 5,113 observations, amounting to a total of 25,565 data points. The dataset is organized as a time series. In this study, missing values were simulated

under a Missing Completely at Random (MCAR) framework to enable controlled methodological comparison; however, real-world meteorological datasets often exhibit Missing at Random (MAR) and Missing Not at Random (MNAR) mechanisms, which are influenced by observed environmental conditions or extreme events [16].

Interpolation Methods

Interpolation is a computational process designed to estimate missing or unknown values within a dataset by using known data points [17]. The accuracy of interpolation results is highly dependent on the algorithm employed [18]. This study evaluates and compares four interpolation methods: linear, spline, stineman, and moving average. Linear interpolation estimates new data points by drawing straight lines between two known values. This method is particularly useful for its simplicity and computational efficiency [19]. In the linear interpolation method, x denotes the independent variable representing the position of the data point (e.g., time index), while x_i and x_{i+1} represent the independent variable values at the i-th and (i+1)-th observed data points, respectively. The variable y denotes the dependent variable corresponding to a given x, with y_i and y_{i+1} representing the observed values at x_i and x_{i+1}. Linear interpolation estimates the value of y at a point x lying between x_i and x_{i+1} by assuming a linear relationship between the two known data points (x_i,y_i) and (x_{i+1}, y_{i+1}). The following formula is employed to calculate the value of y at a specific point x for $i = 1,2,3,\dots,n$:

$$y = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) \tag{1}$$

In the spline interpolation method, x denotes the independent variable representing the position of the data point (e.g., time index), while $x_i, x_{i+1}, x_{i+2}, x_{i+3}$ represent the independent variable values at consecutive observed data points used to construct each spline segment. The variable y denotes the dependent variable corresponding to a given$x$, with $y_i, y_{i+1}, y_{i+2}, y_{i+3}$ representing the observed values at the respective points. Spline interpolation constructs a smooth piecewise polynomial function that passes through the observed data while maintaining continuity and smoothness of the first and/or second derivatives across adjacent intervals. More generally, spline interpolation approximates a function $f(x)$ over the interval $a \le x \le b$ using a spline function $g(x)$, which is defined by partitioning the domain into sub-intervals $a = x_1 < x_2 < \cdots < x_n = b$, where each segment is represented by a polynomial function with coefficients (e.g., b_1 and b_2) that control the local shape and smoothness of the curve [20]. The function $g(x)$, is expressed using the following formula for $i = 1,2,3,\dots,n$:

$$y = y_i + \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) + b_1(x - x_i)(x - x_{i+1}) + b_2(x - x_i)(x - x_{i+1})(x - x_{i+2}) \tag{2}$$

where is

$$b_1 = \frac{\left(\frac{y_{i+2} - y_{i+1}}{x_{i+2} - x_{i+1}}\right) - \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right)}{x_{i+2} - x_i} \tag{3}$$

$$b_2 = \frac{\left[\frac{\left(\frac{y_{i+3}-y_{i+1}}{x_{i+3}-x_{i+2}}\right)-\left(\frac{y_{i+2}-y_{i+1}}{x_{i+1}-x_i}\right)}{x_{i+3}-x_i}\right]-\left[\frac{\left(\frac{y_{i+n+1}-y_{i+n}}{x_{i+2}-x_{i+1}}\right)-\left(\frac{y_{i+n}-y_i}{x_{i+1}-x_i}\right)}{x_{i+3}-x_i}\right]}{x_{i+3}-x_i} \quad (4)$$

Stineman interpolation is a method used to estimate missing values, particularly in datasets with fluctuating or intersecting trends, by employing rational approximations that preserve data continuity and minimize oscillations [21]. In this method, x denotes the independent variable representing the position of the data point (e.g., time index), while x_i and x_{i+1} represent two consecutive observed points that bound the interpolation interval. The variable y denotes the dependent variable at a given x, with y_i representing the observed value at x_i, and y_0 representing a reference or initial value depending on the formulation. The quantities Δy_i and Δy_{i+1} represent successive differences between observed values used to estimate local slopes, while S_i denotes the slope or gradient at point x_i, calculated based on neighboring differences to ensure shape-preserving interpolation. For i =1, 2, 3 , ... , n, the interpolation is computed using different formulations depending on the sign of Δy_i and Δy_{i+1}, allowing the method to adapt to local data behavior and maintain stability. The following formulas, Stineman interpolation formulas is:

$$y = y_0 + \frac{\Delta y_i \Delta y_{i+1}}{\Delta y_i + \Delta y_{i+1}}, \qquad \text{If } \Delta y_i, \Delta y_{i+1} > 0 \quad (5)$$

$$y = y_0 + \frac{\Delta y_i \Delta y_{i+1}(x - x_i + x - x_{i+1})}{(\Delta y_i + \Delta y_{i+1})(x_{i+1} - x_i)}, \qquad \text{If } \Delta y_i, \Delta y_{i+1} > 0 \quad (6)$$

where is

$$\Delta y_i = y_i + S_i(x - x_i) - y_0 \quad (7)$$

$$\Delta y_{i+1} = y_{i+1} + S_i(x - x_{i+1}) - y_0 \quad (8)$$

$$S_i = \frac{x_{i+1} - x_i}{y_{i+1} - y_i} \quad (9)$$

The last, moving average interpolation method estimates missing values by replacing a missing observation at time *t* with the average of k preceding observations [22]. In this method, y denotes the value of the dependent variable at a given time point, while y_i represents the observed value at time index i. The term y_{i-2} represents a prior observation (two time steps before i), and y_0 denotes an initial or boundary value depending on the formulation. The parameter k represents the moving average window size, defined as the number of observations included in the averaging process. For i =1, 2, 3, ... , n, the interpolated value is computed as the mean of the previous k observed values, allowing the method to smooth short-term fluctuations while preserving the overall trend of the data. The choice of k influences the balance between smoothing and responsiveness, where smaller values retain local

variability and larger values produce smoother estimates. The moving average interpolation method is expressed as:

$$y = \frac{(y_i + y_{i-2} + \cdots + y_0)}{k} \tag{10}$$

Performance Evaluation Metric

Performance evaluation metrics are utilized to measure the effectiveness of each interpolation method in achieving the desired level of accuracy. This study employs three primary metrics: MAE, RMSE), and MASE. In Equations (11)–(13), the variables are defined as follows: y_t denotes the observed (actual) value at time t, while $\hat{y}_t$ represents the interpolated (estimated) value at time t, and n is the total number of observations used in the evaluation. The index t indicates the time step, with t = 1, 2, ... , n.

The MAE measures the average magnitude of absolute differences between observed and estimated values, providing a straightforward measure of accuracy. The RMSE measures the square root of the average squared differences, giving greater weight to larger errors. The MASE scales the MAE by the mean absolute difference between consecutive observations, $|y_t - y_{t-1}|$, providing a scale-independent measure that allows comparison across different datasets. In the MASE formulation, the denominator represents the mean absolute error of a naïve one-step forecast, which serves as a benchmark for scaling. These metrics provide a quantitative basis for assessing the reliability of the interpolation methods. The formulas for MAE, RMSE, and MASE are presented in Equations, as outlined by [23, 24].

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{12}$$

$$MASE = \frac{MAE}{\frac{1}{n-1}\sum_{i=2}^{n}|y_t - y_{t-1}|} \tag{13}$$

Stage of Data Analysis

The analysis is to determine the optimal interpolation method that is done in six stages. The study first calculated descriptive statistics, such as the minimum, mean, and maximum values as measures of central tendency. It aims to summarize a dataset by identifying a central or typical value that represents the entire distribution of data [25]. Additionally, it calculated the standard deviation and interquartile range as measures of dispersion to understand the variability of a set of initial data used in the study as a “baseline dataset” [26]. Next, the sum of the missing values for every variable will be computed. Descriptive statistical analysis allows for the identification of patterns and trends within the data [27].

In the second stage, the baseline dataset containing missing values was processed using the interpolation methods: linear (Equation 1), spline (Equation 2), stineman (Equations 5 and 6), and moving average (Equation 10). The interpolated data serves as an "actual dataset" for subsequent analyses. These datasets were further analyzed using descriptive statistics to assess central tendencies and dispersion. The third stage involved introducing simple random deletions to the actual dataset to simulate data loss at three levels: 10% (511 observations), 20% (1,023 observations), and 30% (1,534 observations) from a total of 5,113 observations for each variable. This approach was used to test the robustness and reliability of the interpolation methods while mimicking real-world scenarios of missing data [28].

In the fourth stage, the dataset with simulated missingness was re-interpolated for each percentage of data loss using the same interpolation equations as in the second step. This re-interpolation allowed for an evaluation of how each method responds to varying levels of data incompleteness, creating "prediction dataset". The fifth stage focused on evaluating the performance of the interpolation methods. Performance metrics, including MAE, RMSE, and MASE, were calculated by comparing the actual dataset with the prediction dataset for each variable at different levels of missingness.

In the last stage, the average MAE, RMSE, and MASE values are computed from the comparison results of the five variables with identical percentages of missing data for each interpolation method. The performance metrics are chosen based on assessing data interpolation accuracy, emphasizing both absolute and relative error measurements. Ultimately, this study focuses on identifying the most effective interpolation method by selecting the method with the lowest average MAE, RMSE, and MASE. All stages of analysis in this study were carried out using R-4.4.2 software. R software is free and open-source, licensed by the GNU Project, and distributed under the GNU General Public License [29]. For reproducibility, we set the global random seed using set.seed(141). The use of a fixed seed ensures that missingness patterns and interpolation comparisons are exactly reproducible by independent researchers.

Results and Discussion

Descriptive Statistics for Baseline Data

Descriptive statistics, outlined in Table 1, provide a summarized overview of the primary characteristics of North Aceh's meteorological data from 2010 to 2023. Table 1 highlights the baseline dataset's key statistical measures. Temperature displayed a stable trend, with values ranging between 23.3 °C and 30.3 °C, averaging 26.7 °C. Its low standard deviation of 0.911 suggests minimal fluctuation, indicating consistent behavior over time. Humidity exhibited higher variability, with values ranging from 57.0% to 100.0% and an average of 83.99%. Its standard deviation of 4.315 indicates moderate variability. Rainfall showed the most significant variability, with values spanning from 0.0 mm to 181.7 mm, a mean of 5.6 mm, and a substantial standard deviation of 14.22. These statistics suggest occasional extreme values within the dataset. Sunshine duration averaged 5.57 hours, with a range from 0.0 to 11.7 hours and a moderate standard deviation of 3.007, reflecting a broad distribution. Wind speed proved to be the most consistent variable, with an average of 2.01 m/s, minimal variability

(standard deviation of 0.74), and an interquartile range (IQR) of 0.0, highlighting uniformity in its central tendency. The distribution of each variable is shown in Figure 2. Furthermore, the dataset includes missing values for all meteorological variables analyzed in this study. Rainfall had the highest number of missing observations, totaling 1,484, while wind speed had the fewest missing values, with 38 instances.

Table 1 Summary statistics of baseline dataset.

No	Variable	Min	Mean	Max	IQR	Stdv	Missing Value
1	Temperature	23.30	26.70	30.30	1.20	0.911	38
2	Humidity	57.00	83.99	100.00	6.00	4.315	41
3	Rainfall	0.00	5.60	181.70	4.20	14.220	1,484
4	Sunshine duration	0.00	5.57	11.70	4.80	3.007	281
5	Wind speed	0.00	2.01	5.00	0.00	0.740	33

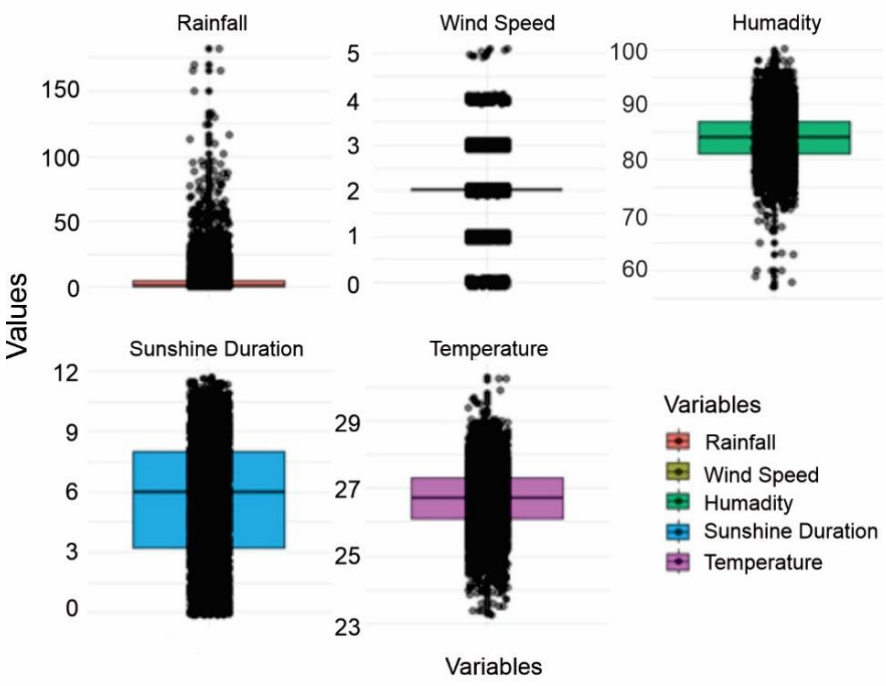

Figure 2 Distribution of meteorological variables.

Furthermore, the intersection size for missing data is shown in Figure 3. It shows distinct patterns of missing data across the five climatic variables in this study. With uncombined missingness, the largest missingness occurs in rainfall (1,381 observations), followed by Sunshine duration (178 observations). Next, joint missingness of sunshine duration and rainfall (77 observations), and a combination of five variables (25 cases). Other combinations are rare (<25 observations). It indicates that missing data are concentrated in rainfall, often coinciding with gaps in rainfall and sunshine duration, likely due to measurement failures during cloudy or rainy conditions. The results of the missingness

suggest a random pattern. Therefore, appropriate handling using interpolation is essential to maintain the reliability and accuracy of subsequent statistical or climatological analyses.

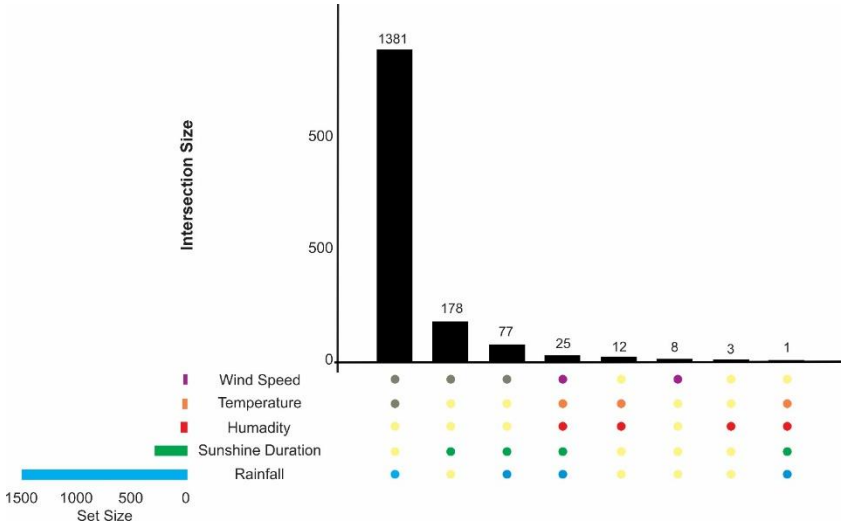

Figure 3 The intersection size for missing data from five variables.

Descriptive Statistics for Actual Dataset

The baseline meteorological data, which initially contained missing values, was completed using various interpolation methods. After filling the missing values, a descriptive analysis was conducted on the resulting dataset, with the summary statistics presented in Table 2. The descriptive statistics for temperature indicate a consistent mean of 26.7 °C across all interpolation methods, with minimal variability reflected by a standard deviation of 0.9 and an interquartile range (IQR) of 1.2. This uniformity suggests all methods are equally reliable in handling temperature data, given its stable characteristics over time. The stability in results highlights the absence of significant deviations introduced by the interpolation methods, making them all suitable for datasets with low variability. For humidity, a similar trend of consistency emerges, with a mean of 84%, ranging from 57% to 100%, and a standard deviation of 4.3 across methods. The narrow IQR of 6.0 further underscores the uniformity in distribution. These findings indicate that all interpolation methods adequately handle humidity data, ensuring reliable outcomes with minimal variation between methods.

However, rainfall demonstrates pronounced differences across interpolation methods. Spline interpolation produces a much higher mean value of 45.3 mm, accompanied by extreme variability, as shown by a maximum value of 838.5 mm and a standard deviation of 152.5. In contrast, stineman and moving average interpolation methods yield significantly lower means (around 6.6 mm and 6.0 mm, respectively) and reduced variability, with standard deviations near 14.2 and 14.3. Linear interpolation, while maintaining a lower mean of 6.1 mm, shows a slightly smaller standard deviation of 13.9. These differences highlight spline interpolation's susceptibility to amplifying variability, making it less suitable for high-variance data like rainfall. For sunshine duration, the mean value of 5.6 hours remains consistent across all methods, with the standard deviation in the same value of 3.0. However, all interpolation

methods except moving average interpolation introduce slightly higher maximum value, as evidenced by its maximum value of 14.1 hours compared to 11.7 hours for the other methods. This information suggests that spline interpolation may slightly exaggerate extremes in the data. Finally, wind speed exhibits exceptional consistency across all methods, with a mean of 2.0 m/s, an IQR of 0.0, and a standard deviation of 0.7. The lack of variation across interpolation methods indicates that all methods are equally effective for handling this variable, which inherently demonstrates minimal variability.

Table 2 Summary statistics of the actual dataset.

No.	Methods	Variable	Min	Mean	Max	IQR	Stdv
1	Linear	Temperature	23.3	26.7	30.3	1.2	0.9
		Humidity	57.0	84.0	100.0	6.0	4.3
		Rainfall	0.0	6.1	181.7	5.9	13.9
		Sunshine duration	0.0	5.6	11.7	4.7	3.0
		Wind speed	0.0	2.0	5.0	0.0	0.7
2	Spline	Temperature	23.3	26.7	30.3	1.2	0.9
		Humidity	57.0	84.0	100.0	6.0	4.3
		Rainfall	0.0	45.3	838.5	11.4	152.5
		Sunshine duration	0.0	5.6	14.1	4.8	3.0
		Wind speed	0.0	2.0	5.0	0.0	0.7
3	Stineman	Temperature	23.3	26.7	30.3	1.2	0.9
		Humidity	57.0	84.0	100.0	6.0	4.3
		Rainfall	0.0	6.6	181.7	7.8	14.2
		Sunshine duration	0.0	5.6	11.7	4.7	3.0
		Wind speed	0.0	2.0	5.0	0.0	0.7
4	Moving average	Temperature	23.3	26.7	30.3	1.2	0.9
		Humidity	57.0	84.0	100.0	6.0	4.3
		Rainfall	0.0	6.0	181.7	5.7	14.3
		Sunshine duration	0.0	5.6	11.7	4.7	3.0
		Wind speed	0.0	2.0	5.0	0.0	0.7

Descriptive Statistics for Prediction Dataset Based on Percentage of Missing Data

Following the removal of observations at predefined levels of missing data, the interpolated data for each variable was analyzed. This step aimed to evaluate how different interpolation methods responded to varying degrees of data incompleteness. The prediction datasets generated through interpolation were subjected to descriptive statistical analysis to identify patterns and trends for each variable, as summarized in Table 3. The summary statistics presented in Table 3 provide a detailed examination of how different interpolation methods give impact the prediction dataset at three different

data thresholds (10%, 20%, and 30%). Across the four interpolation methods: linear, spline, stineman, and moving average, each variable exhibits distinct patterns of predictability and variability.

Temperature shows remarkable consistency across almost all methods, with a narrow range from 23.3 °C to 30.3 °C and stable mean values (26.7 °C). The low standard deviations (0.8 to 0.9) and minimal fluctuations in IQR suggest that temperature is highly predictable, regardless of the interpolation method or data percentage. This stability makes Temperature a reliable variable for forecasting, where even simpler models like linear interpolation are sufficient. In contrast, Humidity demonstrates higher variability, particularly in the Spline interpolation method, where maximum values (e.g., 105.3%) suggest potential overestimation or anomalies in the data. Despite this, the mean humidity remains around 84%, indicating general stability, though the larger standard deviations point to greater uncertainty in predictions for this variable. The variability is slightly reduced when the data percentage increases, but the fundamental instability remains, particularly in methods like spline.

The most significant variation occurs in rainfall, which exhibits the widest range (0.0 to 838.5 mm). While the mean rainfall is low (around 6.0 to 6.6 mm), the Spline interpolation method leads to extreme outliers, with values reaching 838.5 mm. This high variability is reflected in large standard deviations and IQRs, indicating that Rainfall is inherently difficult to predict accurately, particularly when extreme events are involved. The choice of interpolation method becomes critical here, as more complex models like spline may capture non-linear patterns but also introduce significant uncertainty.

Table 3 Summary statistics of prediction data.

No	Methods	Percentage	Variable	Min	Mean	Max	IQR	Stdv
1	Linear	10%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	58.0	84.0	100.0	6.0	4.2
			Rainfall	0.0	6.2	181.7	6.0	13.7
			Sunshine duration	0.0	5.5	11.7	4.7	2.9
			Wind speed	0.0	2.0	5.0	0.0	0.7
		20%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	100.0	6.0	4.2
			Rainfall	0.0	6.1	181.7	6.1	13.2
			Sunshine duration	0.0	5.5	11.7	4.5	2.9
			Wind speed	0.0	2.0	5.0	0.0	0.7
		30%	Temperature	23.3	26.7	30.3	1.2	0.8
			Humidity	59.0	84.0	100.0	5.2	4.1
			Rainfall	0.0	6.1	181.7	6.2	13.1
			Sunshine duration	0.0	5.6	11.7	4.4	2.8
			Wind speed	0.0	2.0	5.0	0.0	0.7

Table 3 Summary statistics of prediction data ***(cont.)***.

No	Methods	Percentage	Variable	Min	Mean	Max	IQR	Stdv
2	Spline	10%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	58.0	84.0	100.0	6.0	4.3
			Rainfall	0.0	45.6	838.5	12.0	152.5
			Sunshine duration	0.0	5.6	14.1	4.9	3.0
			Wind speed	0.0	2.0	5.4	0.0	0.7
		20%	Temperature	22.6	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	105.1	6.0	4.4
			Rainfall	0.0	45.7	838.5	12.4	152.4
			Sunshine duration	0.0	5.6	19.5	4.8	3.1
			Wind speed	0.0	2.0	5.0	0.0	0.8
		30%	Temperature	22.6	26.7	30.3	1.3	0.9
			Humidity	59.0	84.0	105.3	6.0	4.5
			Rainfall	0.0	45.9	838.5	12.4	152.4
			Sunshine duration	0.0	5.7	19.5	4.7	3.1
			Wind speed	0.0	2.0	6.7	0.0	0.8
3	Stineman	10%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	58.0	84.0	100.0	6.0	4.3
			Rainfall	0.0	6.6	181.7	8.0	13.9
			Sunshine duration	0.0	5.5	11.7	4.7	2.9
			Wind speed	0.0	2.0	5.0	0.0	0.7
		20%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	100.0	6.0	4.2
			Rainfall	0.0	6.6	181.7	8.2	13.5
			Sunshine duration	0.0	5.5	11.7	4.5	2.9
			Wind speed	0.0	2.0	5.0	0.0	0.7
		30%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	100.0	5.3	4.1
			Rainfall	0.0	6.5	181.7	8.1	13.4
			Sunshine duration	0.0	5.6	11.7	4.4	2.8
			Wind speed	0.0	2.0	5.0	0.0	0.7
4	Moving average	10%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	58.0	84.0	100.0	6.0	4.2
			Rainfall	0.0	6.6	181.7	8.0	13.8
			Sunshine duration	0.0	5.5	11.7	4.5	2.9
			Wind speed	0.0	2.0	5.0	0.0	0.7

Table 3 Summary statistics of prediction data ***(cont.)***.

No	Methods	Percentage	Variable	Min	Mean	Max	IQR	Stdv
4	Moving average	20%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	100.0	5.1	4.2
			Rainfall	0.0	6.6	181.7	8.3	13.3
			Sunshine duration	0.0	5.5	11.7	4.3	2.8
			Wind speed	0.0	2.0	5.0	0.0	0.7
		30%	Temperature	23.3	26.7	30.3	1.2	0.9
			Humidity	59.0	84.0	100.0	4.7	4.0
			Rainfall	0.0	6.6	181.7	8.3	13.1
			Sunshine duration	0.0	5.6	11.7	4.1	2.7
			Wind speed	0.0	2.0	5.0	0.0	0.7

For Sunshine duration, the data shows moderate variability, with a range from 0.0 to 19.5 hours. While the mean values are fairly consistent (5.5 to 5.7 hours), the IQR and standard deviations suggest some fluctuations, particularly at higher data percentages. The variation is less extreme compared to Rainfall and Humidity, but still notable enough to influence the precision of forecasting. Wind speed is the most predictable variable, with values consistently ranging from 0.0 to 5.4 m/s and minimal variation in both standard deviation and IQR. This stability indicates that Wind speed can be reliably predicted using any interpolation method, with little impact from data percentage changes.

When comparing the interpolation methods, linear interpolation generally provides the most stable predictions, especially for Temperature and Wind speed, with low variability across all variables. This makes it a strong choice for models requiring consistent results. In contrast, the spline method introduces greater variability, especially for rainfall, where extreme values cause larger standard deviations. While spline interpolation may capture more nuanced patterns in complex datasets, its increased sensitivity to outliers suggests that it may not always provide the most reliable predictions for highly variable variables like Rainfall and Humidity. The stineman and moving average methods show similar performance, offering moderate variability compared to linear interpolation but without the extremes seen in spline. These methods may strike a balance, providing a smoother approach that can be useful when moderate fluctuations are expected, such as in Sunshine duration and Wind speed predictions.

From an operational forecasting perspective, the implications are clear. For stable variables like Temperature and Wind speed, simpler methods like linear interpolation are both effective and efficient, as they provide accurate predictions with minimal complexity. However, for more variable and unpredictable data such as Rainfall, Spline interpolation may offer a more detailed representation of complex weather patterns, though care must be taken to address outliers and variability. The stineman and moving average methods offer an intermediate approach suitable for situations where moderate smoothing is needed without the risk of extreme prediction errors.

Increasing the data percentage from 10% to 30% does not significantly impact the predictions for stable variables like Temperature and Wind speed, but it does improve the understanding of more complex variables like Rainfall and Sunshine duration. While this additional data enhances model robustness, particularly for Rainfall, it also introduces more complexity and potential for error, especially with more sensitive methods like Spline.

Evaluation of Performance Metrics

Figure 4 evaluates four interpolation methods: linear, spline, stineman, and moving average across three data percentages (10%, 20%, and 30%) using MAE, RMSE, and MASE. These metrics are essential for evaluating predictive accuracy, with lower values generally indicating better model performance [30].

Linear Interpolation emerged as the most consistently reliable method, exhibiting the lowest MAE, RMSE, and MASE values across all data percentages. For instance, at 10% data missing, linear interpolation achieved an MAE of 0.218, an RMSE of 1.238, and a MASE of 0.170, remaining stable as the data percentage increased. While the error metrics increased slightly with higher data percentages (MAE = 0.458 at 20% and MAE = 0.666 at 30%), it still outperformed the other methods. These results highlight linear Interpolation as the most efficient and stable choice for predictive modeling in scenarios where simplicity and computational efficiency are key.

In contrast, Spline Interpolation showed higher error values, particularly as the data percentage increased. At 10%, the MAE was 0.267, RMSE was 1.473, and MASE was 0.204, which was higher than Linear Interpolation. As the data percentage grew, spline interpolation's performance deteriorated, with MAE reaching 0.851, RMSE at 2.714, and MASE at 0.657 at 30%. This information suggests that while spline interpolation can model complex relationships in the data, its accuracy declines as the dataset expands, likely due to overfitting or an inability to generalize well to larger datasets. Therefore, while spline might be more suitable for datasets with highly non-linear trends, its use requires careful management of data size and complexity to avoid compromising predictive accuracy.

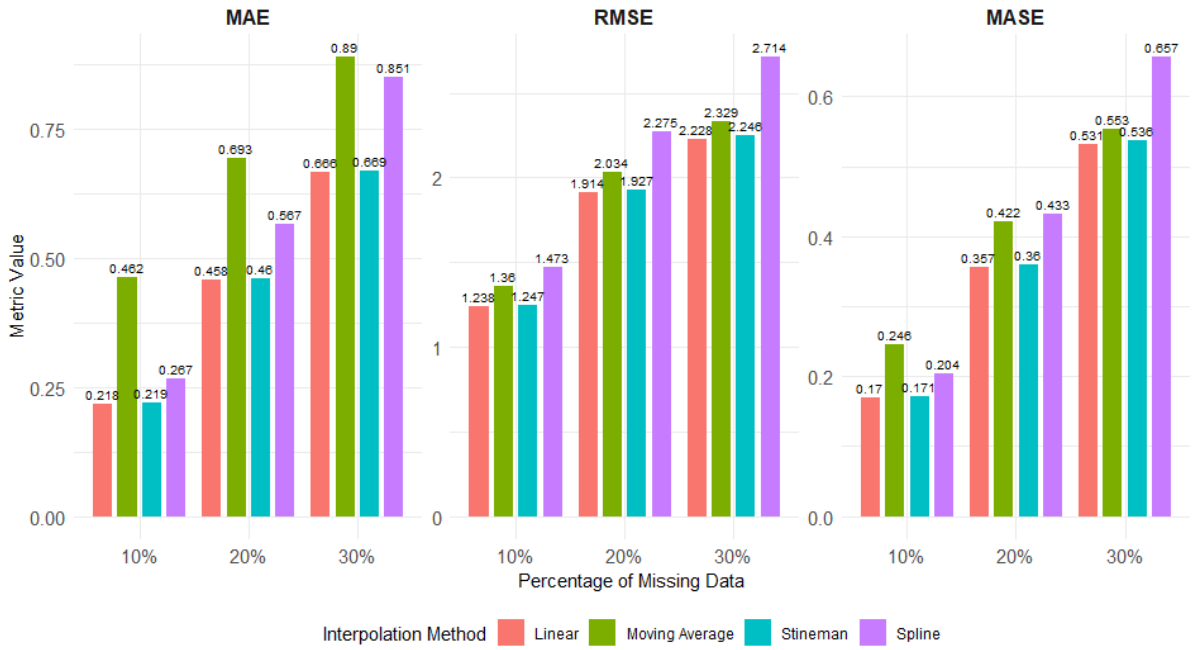

Figure 4 Performance metrics across interpolation methods.

Stineman interpolation exhibited performance similar to linear interpolation, with small increases in error as data percentages rose. At 10%, it had MAE = 0.219, RMSE = 1.247, and MASE = 0.171, closely mirroring linear interpolation. As the data percentage increased, its performance remained relatively stable, with MAE = 0.669 at 30%. This makes stineman interpolation a solid alternative to linear interpolation, providing a balanced approach when small improvements in accuracy are needed without significantly increasing model complexity or computational costs.

On the other hand, moving average interpolation demonstrated the highest error rates across all metrics and data percentages. At 10%, the MAE was 0.462, RMSE was 1.360, and MASE was 0.246, which gradually worsened as the data percentage increased. By 30%, moving average had the highest MAE (0.890), RMSE (2.329), and MASE (0.553), suggesting it was the least accurate method for this dataset. This consistent underperformance implies that while moving average might be effective for smoothing short-term fluctuations, it struggles with capturing the underlying patterns in more complex datasets, especially as the data size increases.

The consistent outperformance of linear interpolation, despite its algorithmic simplicity, provides a compelling argument for its use in long-term tropical meteorological data interpolation. This finding cautiously challenges the assumption that more complex, non-linear methods are inherently superior for time-series data. Our results align with studies like [31], who found linear interpolation effective for filling missing data due to its stability, and contrast with studies that champion spline methods for specific, short-term applications [8, 9]. The discrepancy with spline-based studies can likely be attributed to the long-term nature of our dataset and the inherent characteristics of meteorological variables. Over a 14-year period, the high-frequency fluctuations that spline methods excel at capturing may represent noise rather than signal.

The propensity of spline interpolation to generate extreme values (e.g., rainfall up to 838.5 mm) indicates overfitting, where the model captures random variations in the data rather than the true underlying climatic trend. This is a critical limitation in climate studies where preserving the statistical distribution of variables is paramount. The poor performance of the moving average method underscores its inadequacy for precise data interpolation, confirming its role as a smoothing technique rather than a reliable gap-filling tool [32].

To further investigate the source of the high aggregate errors observed in spline and moving average methods, we analyzed the contribution of each meteorological variable to the overall error metrics. The results indicate that variables with high variability and intermittent behavior, particularly rainfall, contribute disproportionately to the total error. It is due to the presence of abrupt changes and extreme values, which are not well captured by smoothing-based interpolation techniques such as spline and moving average. In contrast, variables with relatively stable temporal patterns, such as temperature and humidity, exhibit lower error values across all interpolation methods. This finding is consistent with previous studies indicating that interpolation performance is highly dependent on the statistical characteristics of the variable, including variance, continuity, and distribution shape [33, 34]. These results suggest that the instability observed in spline interpolation is not uniform across variables but is primarily driven by highly fluctuating time series. Therefore, the selection of interpolation methods should consider variable-specific characteristics, particularly when dealing with non-stationary or extreme-value-dominated data [35, 36].

The generalisability of these findings is strengthened by the study's use of a long-term (2010-2023), high-resolution (daily) dataset encompassing key meteorological variables in tropical regions. While the data is from a specific tropical region in Indonesia, the methodological framework and the comparative analysis of fundamental interpolation algorithms are universally applicable. The results are most directly generalizable to regions with similar climatic regimes (e.g., humid tropics). However, the core finding that simple, stable methods like linear interpolation can be optimal for long-term trend preservation is a significant insight that likely extends to temperate and other climate zones. The study's external validity is contingent on MCAR assumption. For data missing not at random (MNAR), such as systematic sensor failures during extreme weather, the performance of these methods may differ, and the generalisability would need to be re-evaluated.

This research carries important implications for the theoretical framework of time-series interpolation and environmental data science. It provides empirical evidence that challenges the "complexity equals accuracy" paradigm often assumed in numerical analysis. For long-term climatic datasets, where the goal is often to preserve low-frequency trends and central tendencies for climate modeling and analysis, a simpler model can be not only sufficient but optimal. This supports the principle of parsimony (Occam's razor) in the context of geoscientific data curation. Our findings suggest that the theoretical optimality of higher-order polynomials in interpolation does not necessarily translate to practical superiority in real-world, noisy, long-term environmental datasets, thereby refining the theoretical expectations for applying these methods in climatology.

Future research should build upon this study's benchmark of interpolation performance in long-term meteorological data by systematically comparing the demonstrated stability of simple interpolation under MCAR conditions with methods that explicitly capture temporal dependence, spatial correlation, and nonlinear relationships. Promising directions include (i) spatio-temporal kriging and hybrid interpolation approaches that leverage information from neighbouring stations [37, 38], (ii) multiple imputation frameworks that account for uncertainty through repeated data reconstruction [39, 40], and (iii) machine-learning and ensemble techniques, such as LightGBM and stacked models, which have shown strong capability in handling complex environmental data gaps [41, 42]. In particular, hybrid strategies such as combining linear interpolation for low-variance variables with machine learning or kriging for highly volatile variables like rainfall, or modelling interpolation residuals using data-driven methods offer a balanced approach between robustness and adaptability. Additionally, future studies should validate these findings across diverse climatic and geographical settings, extend the evaluation framework to other environmental domains (e.g., hydrology and air quality), and rigorously assess method performance under more realistic missing data mechanisms, including MAR and MNAR. Such comprehensive investigations will help establish practical, context-sensitive guidelines for selecting appropriate imputation techniques in real-world applications.

For practitioners, including meteorologists, climatologists, and data curators, our findings offer a clear and actionable recommendation: linear interpolation should be the default, first-choice method for imputing missing values in long-term daily meteorological time series. Its implementation is straightforward, computationally efficient, and readily available in most statistical software packages [31, 43, 44]. This practice can significantly enhance the quality and homogeneity of climate datasets used for operational forecasting, agricultural planning, and renewable energy assessment. The study also provides a practical warning against the uncritical use of spline interpolation for variables with high variance, such as rainfall, as it can introduce significant biases and artificial extremes that distort subsequent analyses.

At a policy level, reliable data is the cornerstone of evidence-based decision-making for climate adaptation and mitigation strategies, directly supporting the targets of Sustainable Development Goal 13 (Climate Action). This study provides a validated, low-cost methodological standard for improving the quality of national meteorological archives. National bodies, such as BMKG in Indonesia, can integrate these findings into their data management and quality control protocols. By standardizing the use of the most effective interpolation technique, policymakers can be more confident in the climate trends and extremes derived from historical data, leading to more robust national climate risk assessments, disaster preparedness plans, and sustainable agricultural policies. Ensuring data integrity at this fundamental level strengthens the entire chain of evidence used to inform critical climate policy.

Despite the robust findings, this study has several limitations. First, the simulation was based on the MCAR assumption. In reality, missing data in meteorology can be systematic MNAR and MAR also. The direction of potential bias here is unpredictable and would require a separate investigation. Second, the study was conducted on data from a single geographical location. While the methodology

is sound, the absolute performance of each method may vary in regions with fundamentally different climate dynamics (e.g. arid or polar regions). Third, we focused on univariate interpolation methods. While this is common practice, spatiotemporal interpolation methods that incorporate data from neighboring stations could potentially yield better results, though they require more complex data and models. Finally, the moving average method's performance is sensitive to the window size (k), which was fixed in this study; optimizing this parameter for each variable might have altered its relative performance, though it is unlikely to have surpassed linear interpolation.

Conclusions

This study assessed the effectiveness of four interpolation methods; linear, spline, stineman, and moving average, under varying levels of missing data (10%, 20%, and 30%) using three performance metrics: MAE, MASE, and RMSE for long term tropical meteorological data. The results indicate that spline interpolation performs well for modeling complex data relationships, but its accuracy declines as the dataset size increases. While suitable for non-linear trends, its use requires careful management of data complexity to avoid reduced accuracy. Moving average interpolation, although computationally simpler, exhibited the highest error rates across all levels of missing data, making it less suitable for high-precision tasks. It struggled to capture underlying patterns, particularly as data sparsity increased.

In contrast, stineman interpolation offered a balanced solution, yielding modest improvements in accuracy without substantial computational cost. However, linear interpolation consistently outperformed the other methods, demonstrating the lowest error values across all performance metrics and missing data levels. Its efficiency and stability make it the optimal choice for predictive missing value when simplicity and computational efficiency are essential. This robustness makes it well-suited for applications requiring accurate data recovery, even with increasing data loss. Overall, the findings provide a practical framework for selecting interpolation methods based on data characteristics and accuracy requirements. In conclusion, by identifying linear interpolation as the most robust method for handling missing data in long-term meteorological series, this study provides a practical and effective solution for enhancing climate data integrity. The improved data quality directly contributes to the targets of SDG 13 by strengthening the capacity for climate change assessment, planning, and management, thereby supporting actionable climate initiatives.

Acknowledgements

The author would like to thank Universitas Syiah Kuala for supporting and funding this study and publication (Grant Number 509/UN11.2.1/PG.01.03/SPK/PTNBH/2024).

References

1. Marchi M, Castellanos-Acuña D, Hamann A, Wang T, Ray D, Menzel A. ClimateEU, scale-free climate normals, historical time series, and future projections for Europe. Sci Data. 2020;7(428):1–9.

2. More KS, Wolkersdorfer C. Exploring advanced statistical data analysis techniques for interpolating missing observations and detecting anomalies in mining influenced water data. ACS ES&T Water. 2024;4(3):1036–45.
3. Alejo-Sanchez LE, Márquez-Grajales A, Salas-Martínez F, Franco-Arcega A, López-Morales V, Acevedo-Sandoval OA, et al. Missing data imputation of climate time series: A review. MethodsX. 2025;15(1):1–19.
4. Jones RL, Kharb A, Tubeuf S. The untold story of missing data in disaster research: a systematic review of the empirical literature utilising the Emergency Events Database (EM-DAT). Environ Res Lett. 2023;18(10):1–10.
5. Wang Y, Liu X, Liu R, Zhang Z. Research progress on spatiotemporal interpolation methods for meteorological elements. Water. 2024;16(6):1–17.
6. Che X, Zhang HK, Li ZB, Wang Y, Sun Q, Luo D, et al. Linearly interpolating missing values in time series helps little for land cover classification using recurrent or attention networks. ISPRS J Photogramm Remote Sens. 2024;212(1):73–95.
7. Wicher-Dysarz J, Dysarz T, Jaskuła J. Uncertainty in determination of meteorological drought zones based on standardized precipitation index in the territory of Poland. Int J Environ Res Public Health. 2022;19(23):15797.
8. Azizan I, Karim SABA, Suresh Kumar Raju S. Fitting rainfall data by using cubic spline interpolation. Abdul Karim SA, Zainuddin N, Yusof MH, Sa'ad N, editors. MATEC Web Conf. 2018;225:05001.
9. Yasmin AA, Azahra AS, Purwani S. The application of cubic spline in rainfall modelling in Bogor and its impact on paddy production. Commun Math Biol Neurosci. 2024;2024:31.
10. Bleidorn MT, Pinto W de P, Schmidt IM, Mendonça ASF, Reis JAT dos. Methodological approaches for imputing missing data into monthly flows series. Rev Ambient Agua. 2022;17(2):e2795.
11. Avila ML, Alonso AM, Peña D. Modelling time series with multiple seasonalities: an application to hourly NO2 pollution levels. Stoch Environ Res Risk Assess. 2025;39(5):2063–93.
12. Portela MM, Espinosa LA, Zelenakova M. Long-term rainfall trends and their variability in mainland Portugal in the last 106 years. Climate. 2020;8(12):146.
13. Kim T, Ko W, Kim J. Analysis and impact evaluation of missing data imputation in day-ahead PV generation forecasting. Appl Sci. 2019;9(1):204.
14. Anjomshoaa A, Salmanzadeh M. Filling missing meteorological data in heating and cooling seasons separately. Int J Climatol. 2019;39(2):701-10.
15. BPS-Statistics Indonesia. Aceh Province Gross Regional Domestic Product by Business Field Quarter 3 in 2023. Banda Aceh; 2023.
16. Heymans MW, Twisk JWR. Handling missing data in clinical research. J Clin Epidemiol. 2022;151(1):185–8.
17. Mahmoud A, Yuan X, Kheimi M, Yuan Y. Interpolation accuracy of hybrid soft computing techniques in estimating discharge capacity of triangular Labyrinth Weir. IEEE Access.

2021;9(1):6769–85.

18. Pratama HA, Sam'an M. Implementasi metode interpolasi bilinear untuk perbesaran skala citra. J Komput dan Teknol Inf. 2023;1(1):21–5.
19. Kong Q, Siauw T, Bayen AM. Python programming and numerical methods. st ed. Merken S, editor. London: Elsevier; 2021. p. 1–452.
20. Segeth K. Some splines produced by smooth interpolation. Appl Math Comput. 2018;319(1):387–94.
21. Arjasakusuma S, Pratama AP, Lestari I. Assessment of gap-filling interpolation methods for identifying mangrove trends at Segara Anakan in 2015 by using Landsat 8 OLI and Proba-V. Indones J Geogr. 2020;52(3):1–9.
22. Mohamad NB, Lai AC, Lim BH. A case study in the tropical region to evaluate univariate imputation methods for solar irradiance data with different weather types. Sustain Energy Technol Assess. 2022;50:101764.
23. Etukuru RR. AI-driven time series forecasting: complexity-concious prediction and decision-making. Bloomington (IN): iUniverse; 2023.
24. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. 3rd ed. Melbourne: OTexts; 2021.
25. Sofyan H, Diba F, Susanti SS, Marthoenis M, Ichsan I, Sasmita NR. The state of diabetes care and obstacles to better care in Aceh , Indonesia: a mixed - methods study. BMC Health Serv Res. 2023;1:271.
26. Azharuddin, Sasmita NR, Idroes GM, Andid R, Raihan, Fadlilah T, et al. Patient satisfaction and its socio-demographic correlates in Zainoel Abidin hospital, Indonesia: a cross-sectional study. Unnes J Public Heal. 2023;12(2):57–67.
27. Reskiaddin LO, Ahsan A, Fitri A, Hubaybah H, Putri FE, Sasmita NR. Evaluating the impact of smoke-free policies in Jambi, Indonesia: a mixed-methods approach. Asian Pacific J Cancer Prev. 2025;26(5):1815–21.
28. Qiu H, Chen H, Xu B, Liu G, Huang S, Nie H, et al. Multiple types of missing precipitation data filling based on ensemble artificial intelligence models. Water. 2024;16(22):3192.
29. Saputra A, Sofyan H, Kesuma ZM, Sasmita NR, Wichaidit W, Chongsuvivatwong V. Spatial patterns of tuberculosis in Aceh province during the COVID-19 pandemic: a geospatial autocorrelation assessment. IOP Conf Ser Earth Environ Sci. 2024 Jun 1;1356: 012099.
30. Ulhaq MZ, Farid M, Aziza ZI, Nuzullah TMF, Syakir F, Sasmita NR. Integration of machine learning and time series analysis for upwelling prediction dashboard in lake Laut tawar,Indonesia: a study based on climate forecasting. Theor Appl Climatol. 2025;156(9):463.
31. Huang G. Missing data filling method based on linear interpolation and lightgbm. J Phys Conf Ser. 2021;1754:012187.
32. Yaisamut O, Xie S, Charusiri P, Dong J, Wen W. Prediction of Au-associated minerals in Eastern Thailand based on stream sediment geochemical data analysis by S-A multifractal model. Minerals.

2023;13(10):1297.

33. Ding Q, Wang Y, Zhuang D. Comparison of the common spatial interpolation methods used to analyze potentially toxic elements surrounding mining regions. J Environ Manage 2018;212:23–31.
34. Lee C. Long-term wind speed interpolation using anisotropic regression kriging with regional heterogeneous terrain and solar insolation in the United States. Energy Rep. 2022;8:12–23.
35. Lepot M, Aubin JB, Clemens F. Interpolation in time series: an introductive overview of existing methods, their performance criteria and uncertainty assessment. Water. 2017;9(10):796.
36. Nag P, Sun Y, Reich BJ. Spatio-temporal DeepKriging for interpolation and probabilistic forecasting. arXiv; 2023.
37. Huang J, Lu C, Huang D, Qin Y, Xin F, Sheng H. A spatial interpolation method for meteorological data based on a Hybrid Kriging and machine learning approach. Int J Climatol. 2024;44(15):5371–80.
38. Liu H, Cai C, Li P, Wang Y, Zhao M, Tang C, et al. Hybrid prediction system for reliable multi-seasonal sustainable energy generation under meteorological and environmental volatility. Sci Rep. 2026;16:8637.
39. Chapon A, Ouarda TBMJ, Hamdi Y. Imputation of missing values in environmental time series by D-vine copulas. Weather Clim Extrem. 2023;41:100591.
40. Hua V, Nguyen T, Dao MS, Nguyen HD, Nguyen BT. The impact of data imputation on air quality prediction problem. Gill AR, editor. PLoS One. 2024;19(9):e0306303.
41. Lalic B, Stapleton A, Vergauwen T, Caluwaerts S, Eichelmann E, Roantree M. A comparative analysis of machine learning approaches to gap filling meteorological datasets. Environ Earth Sci. 2024;83(24):679.
42. Pastorini M, RodrÍguez R, Etcheverry L, Castro A, Gorgoglione A. Enhancing environmental data imputation: a physically-constrained machine learning framework. Sci Total Environ. 2024;926:171773.
43. Jung M, Kim DY. DEFT: a dynamic environmental filtering and thresholding algorithm for adaptive headlamp control using ride height sensors. Electronics. 2024 Dec 4;13(23):4788.
44. Wang J. Data interpolation methods with the UNet-based model for weather forecast. Int J Data Sci Anal. 2024;20:2025-38.