JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 8, No. 2. Maret 2026 DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

A SEASONAL IMPUTATION METHOD FOR ADDRESSING MISSING DATA IN ENVIRONMENTAL IOT SENSOR TIME SERIES

Ardiansyah Ramadhan[1], Surya Michrandi Nasution[2], Reza Rendian Septiawan[3], I Kadek Nuary Trisnawan[4], Angel Metanosa Afinda[5]

Computer Engineering Study Program
Telkom University
ardiansyahramadhanar@telkomuniversity.ac.id[1], michrandi@telkomuniversity.ac.id[2], ,zaseptiawan@telkomuniversity.ac.id[3], ikadeknuarytrisnawan@telkomuniversity.ac.id[4], angelmetanosa@telkomuniversity.ac.id[5]

Abstract
Missing and incomplete observations in Environmental IoT sensor networks reduce data reliability and disrupt analyses, especially for temperature and humidity time series exhibiting strong diurnal seasonality. This study develops and evaluates a seasonal imputation method to address missing data in IoT-based environmental monitoring, using a workflow of anomaly detection, outlier removal, time-of-day-aware imputation, and performance evaluation under varying missing-rate scenarios. Key challenges include sensor noise, connectivity issues, and intermittent hardware failures, which degrade data integrity and affect trend analysis, forecasting, and anomaly detection. To mitigate these, the method uses hourly and minute-level seasonal patterns after filtering out physically unrealistic values. Experimental results show high accuracy and robustness in reconstructing temperature and humidity data: temperature imputation achieves MAE values of approximately 0.86–0.87°C, and humidity yields MAE values of 3.92–4.01%RH, with no performance drop even at 50% data loss. The imputed series preserves natural diurnal dynamics without introducing distortions, effectively restoring continuity and structural consistency in environmental IoT time series for reliable modeling, feature extraction, and decision support.

Keywords: Internet of Things; Seasonal imputation; Incomplete data; Time-series analysis;

Abstrak
Data yang hilang dan tidak lengkap pada jaringan sensor Environmental Internet of Things (IoT) dapat menurunkan keandalan data dan mengganggu proses analisis, terutama pada deret waktu suhu dan kelembapan yang memiliki pola musiman harian (diurnal) yang kuat. Penelitian ini mengembangkan dan mengevaluasi metode imputasi musiman untuk menangani data yang hilang dalam pemantauan lingkungan berbasis IoT, dengan alur kerja yang mencakup deteksi anomali, penghapusan outlier, imputasi berbasis waktu (time-of-day-aware), serta evaluasi kinerja pada berbagai skenario tingkat kehilangan data. Tantangan utama yang dihadapi meliputi gangguan sensor, masalah konektivitas, dan kegagalan perangkat keras yang bersifat sementara, yang dapat menurunkan integritas data serta memengaruhi analisis tren, peramalan, dan deteksi anomali. Untuk mengatasi hal tersebut, metode ini memanfaatkan pola musiman pada tingkat jam dan menit setelah terlebih dahulu menyaring nilai-nilai yang secara fisik tidak realistis. Hasil eksperimen menunjukkan akurasi dan ketahanan yang tinggi dalam merekonstruksi data suhu dan kelembapan, dengan nilai MAE suhu sekitar 0,86–0,87°C dan MAE kelembapan sebesar 3,92–4,01%RH, tanpa penurunan kinerja bahkan pada kondisi kehilangan data hingga 50%. Deret waktu hasil imputasi tetap mempertahankan dinamika diurnal alami tanpa menimbulkan distorsi, sehingga mampu memulihkan kontinuitas dan konsistensi struktural pada deret waktu IoT lingkungan untuk mendukung pemodelan, ekstraksi fitur, dan sistem pendukung pengambilan keputusan yang andal.

Kata kunci: Internet of Things; Imputasi musiman; Data tidak lengkap; Analisis deret waktu;

INTRODUCTION

A diverse array of imputation techniques has been developed to address missing data, encompassing fundamental statistical methods as well as sophisticated machine learning and time-series forecasting models (Ramadhan, Indrawati, et al., 2025; Seu et al., 2022). Although intricate methodologies may enhance precision, they often require substantial computational resources and are not always practical for resource-constrained IoT implementations (Hudda & Haribabu, 2025). In fact, many environmental sensor datasets exhibit smooth, periodic patterns, particularly daily cycles, making seasonal or time-of-day-based imputation an effective and suitable option (Fatyanosa et al., 2023).

Despite this potential, limited studies have systematically evaluated lightweight seasonal imputation strategies using real-world environmental IoT datasets. To address this gap, this study investigates the effectiveness of seasonal imputation techniques for handling missing values in environmental IoT sensor time series. The dataset consists of temperature and humidity measurements collected over a one-month period using a DHT22 sensor deployed in the Greater Bandung Basin. Although the sensor data were transmitted automatically through an IoT infrastructure, intermittent communication losses resulted in substantial missing observations in the time series.

This study implements and evaluates two seasonal-based imputation approaches—time-of-day averaging (hourly seasonal imputation) and global median imputation—to reconstruct the missing data. The objective is to identify a reliable and computationally efficient imputation strategy capable of restoring missing sensor values while preserving the underlying temporal patterns of environmental variables. The findings demonstrate that lightweight seasonal imputation methods can effectively reconstruct missing observations in real-world IoT datasets characterized by strong periodicity. Consequently, this study provides practical insights for developers and operators of environmental monitoring systems seeking to improve data continuity without incurring significant computational overhead. In addition, the results contribute to the broader understanding of efficient data preprocessing techniques for environmental IoT analytics.

.

RESEARCH METHODS

The suggested methodology comprises two primary stages: (1) a pre-imputed seasonal technique for the preparation of the IoT dataset, and (2) a post-imputed seasonal technique for the use and assessment of imputation methods. The comprehensive procedure is depicted in Figure 1.

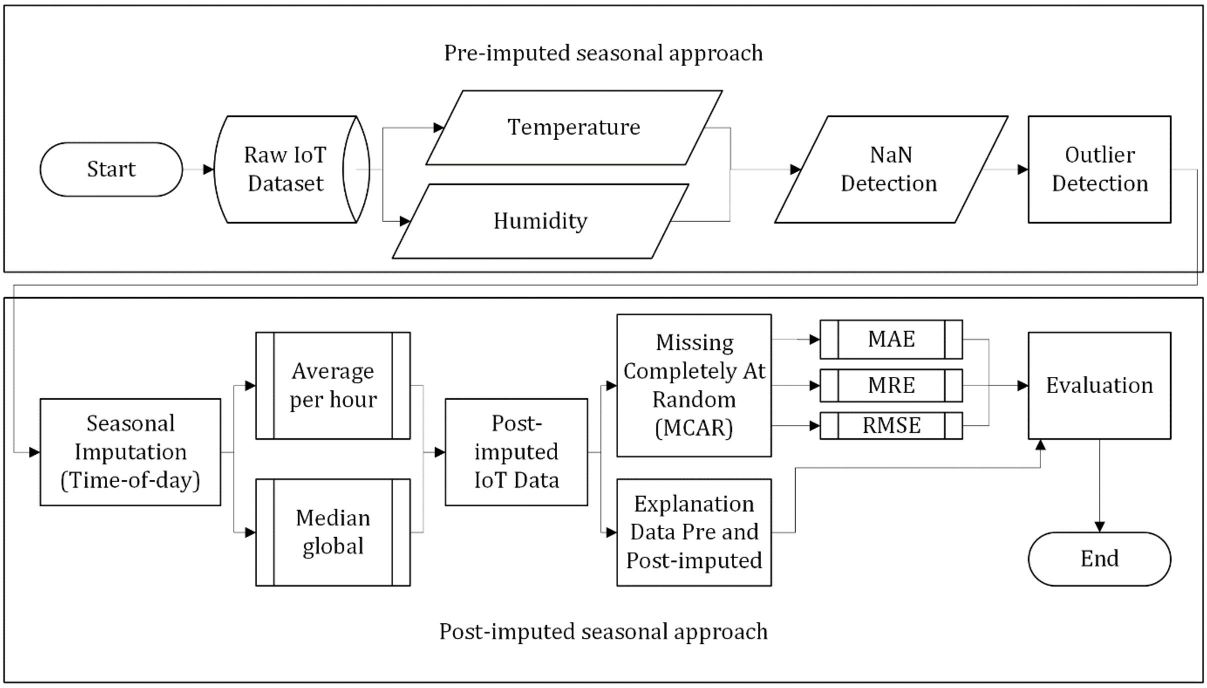

Figure 1. Pre- and Post-Imputed Seasonal Methodology

In the seasonal imputation process, missing values are estimated by exploiting the daily periodic patterns commonly observed in environmental IoT sensor data. Specifically, when a value at time t is missing, it is replaced by the average of historical observations recorded at

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

the same hour of the day. For example, if the temperature measurement at 14:00 on a given day is missing, the imputed value is calculated as the mean of all available temperature observations recorded at 14:00 across previous days. This approach allows the imputation method to capture diurnal seasonal behavior in environmental variables such as temperature and humidity.

A. Pre-Imputed Seasonal Approach
1) Identification of Missing Values

This study identifies missing data by analyzing discrepancies between the timestamps of recorded measurements and the timestamps expected based on the system's sampling period (Seu et al., 2022; Shadbahr et al., 2023; Zhao & Udell, 2024). Let the time-indexed IoT dataset be denoted by equation 1.

$$X = \{x_t \mid t = 1,2,\dots,T\} \tag{1}$$

where each element x_t denotes the sensor measurement at time t. A data point is classified as missing when no observation is recorded at the expected timestamp, which can be formally expressed as equation 2.

$$x_t = \text{NaN} \quad \text{if no observation is recorded at time t.} \tag{2}$$

Identification of missing values is performed by correlating the actual measurement times with a theoretical reference time grid generated from the specified sample frequency (Fatyanosa et al., 2024; Mohammed et al., 2021). Any timestamp without an associated measurement is considered a missing observation, serving as the foundation for subsequent imputation methods designed to restore the continuity and integrity of the IoT time series (Ramadhan, Indrawati, et al., 2025).

2) Outlier Detection

Outliers in the dataset are detected using the Interquartile Range (IQR) criterion, which provides a robust statistical approach for identifying extreme deviations in sensor readings (Bhattacharya et al., 2023; Boukerche et al., 2020; Muhr et al., 2023; Smiti, 2020). Let Q_1 and Q_3 denote the first and third quartiles of the observed values, respectively, and IQR = Q_3-Q_1. A data point x_t is classified as an outlier if it lies outside the threshold bounds defined by Tukey's rule, expressed as equation 3.

$$\text{Outlier } (x_t) = \begin{cases} 1, x_t < Q_1 - 1.5\,\text{IQR} \\ 1, x_t > Q_3 + 1.5\,\text{IQR} \\ 0, otherwise \end{cases} \tag{3}$$

All identified outliers are subsequently removed or reassigned as NaN to prevent them from biasing the imputation process and to ensure that the reconstruction of missing values is based solely on reliable observations (Mahajan et al., 2020; Saleem et al., 2021; Seo, 2006).

B. Post-Imputation Seasonal Method

To reconstruct missing values in the IoT sensor time series, two imputation methods are employed and subsequently evaluated under a controlled missingness scenario (Mara Ribeiro & Leite de Castro, 2022; Qin et al., 2021). The first method, Seasonal Imputation (Time-of-Day / Hourly Average), estimates a missing value by leveraging the daily periodicity commonly observed in environmental variables. Let $h(t)$ denote the hour extracted from timestamp t (Kwok et al., 2023; Mara Ribeiro & Leite de Castro, 2022; Qin et al., 2021; Vasenin et al., 2024). For each missing observation, the imputed value $\hat{x}_t$ is computed as the average of all historical measurements recorded at the same hour of the day, expressed as equation 4.

$$\hat{x}_t = \frac{1}{N_{h(t)}} \sum_{i \in H(t)} x_i \tag{4}$$

where $H(t)$ represents the set of all timestamps sharing the same hour as t, and $N_{h(t)}$ denotes the number of available observations in that set. This method effectively captures diurnal seasonal patterns and provides a computationally efficient imputation strategy suitable for resource-constrained IoT deployments. (Mara Ribeiro & Leite de Castro, 2022; Vasenin et al., 2024)

As a baseline, Global Median Imputation is applied by replacing each missing value with the median of all valid observations in the dataset. Formally, like an equation 5.

$$\hat{x}_t = \text{median}(X_{\text{obs}}), \tag{5}$$

Where $X_{\text{obs}} = \{x_t \mid x_t \neq \text{NaN}\}$. denotes the set of all observed (non-missing) sensor readings. This approach provides a stable reference for evaluating the added value of incorporating seasonal information.

To objectively assess the accuracy of both imputation methods, a Missing Completely at Random (MCAR) mechanism is simulated by intentionally masking a subset of complete observations (Kwok et al., 2023; Lee & Charles Huber, 2021; Pham et al., 2022). For each selected timestamp t, the true value is replaced with a missing indicator such that allowing the reconstructed value $\hat{x}_t$ to be directly compared against its original counterpart during the evaluation phase as shown in equation 6.

$$x_t^{\text{masked}} = \text{NaN}, t \in M, \quad (6)$$

Compared with common imputation techniques such as linear interpolation and k-nearest neighbors (KNN) (Juna et al., 2022), seasonal imputation offers several advantages for environmental IoT sensor data. Many environmental variables exhibit strong diurnal patterns, with measurements at the same hour across different days often similar. By leveraging this periodicity, seasonal imputation can produce more contextually relevant estimates than linear interpolation, which assumes a simple linear transition between adjacent observations. Additionally, seasonal imputation is computationally efficient, requiring only the aggregation of historical values at the same hour, whereas KNN requires distance calculations and neighbor searches, which increase computational cost. Furthermore, seasonal imputation is more robust to longer missing intervals, since it relies on historical seasonal patterns rather than nearby observations. Therefore, it provides a practical and efficient approach for handling missing data in environmental IoT time series.

C. Evaluation Metrics

To quantitatively assess the accuracy of the imputation methods, three standard error metrics are employed: Mean Absolute Error (MAE), Mean Relative Error (MRE), and Root Mean Square Error (RMSE) (Chicco et al., 2021). These metrics compare the reconstructed values $\hat{x}_t$ with the corresponding ground-truth observations x_t within the masked subset of the dataset.

The Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between the true and imputed values, defined as equation 7.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|x_t - \hat{x}_t| \quad (7)$$

The Mean Relative Error (MRE) evaluates the relative deviation by normalizing the absolute error with respect to the true value, expressed as equation 8.

$$\text{MRE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right|. \quad (8)$$

The Root Mean Square Error (RMSE) penalizes larger errors more heavily by computing the square root of the average squared differences between the true and imputed observations equation 9.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (x_t - \hat{x}_t)^2} \quad (9)$$

Collectively, these metrics capture complementary aspects of reconstruction performance—namely absolute deviation, proportional error, and squared error sensitivity. Integrated with the preceding stages of IoT data preprocessing, seasonal time-of-day imputation, baseline statistical imputation, controlled MCAR simulations, and comparative evaluation, this framework enables an effective and computationally efficient assessment of lightweight imputation techniques tailored for environmental IoT deployments (Kwok et al., 2023; Lee & Charles Huber, 2021; Pham et al., 2022).

To quantify the uncertainty of the evaluation metrics, 95% confidence intervals (CI) were calculated based on the mean and standard deviation obtained from repeated missing-data simulations. The CI was computed using the standard formula at equation (10).

$$CI = \bar{x} \pm 1.96\frac{s}{\sqrt{n}} \quad (10)$$

where $\bar{x}$ is the mean error value, s is standard deviation, and n represents the number of experiments. The proposed methodology was implemented in Python using the Pandas and NumPy libraries for time-series processing and statistical analysis. The MCAR masking experiment was repeated 50 times for each missing-rate scenario (1%, 5%, 10%, 20%, and 50%) to ensure statistical robustness. All computations were conducted on a standard workstation environment.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

The repeated simulations enabled the estimation of mean error metrics and their corresponding 95% confidence intervals.

RESULTS AND DISCUSSION

Prior to presenting the comprehensive results, it is imperative to delineate the overall structure and completeness of the temperature and humidity information gathered during the monitoring period, as outlined in Table 1. Comprehending the ratio of valid to missing observations provides essential context for further studies, such as seasonal pattern classification, anomaly detection, and the assessment of the imputation strategy's efficacy. This preliminary overview guarantees that the interpretation of results is based on a precise depiction of the dataset's integrity, temporal continuity, and sensor dependability.

Table 1. Counts of non-null and missing values Internet of Things Unprocessed Dataset

Parameters	Non-null counts	Missing counts
Temperature	120,740	15,886
Humidity	120,738	15,888

The completeness data in Table 1 reveal that both temperature and humidity exhibit comparable percentages of missing values, with roughly 11–12% of the records missing. This degree of missingness is common in long-term IoT deployments and is often caused by packet loss, intermittent sensor outages, connection instability, or power fluctuations. The analogous missing rates for both parameters indicate that the data deficiencies are more likely attributable to systemic issues than to failures of individual sensors. Thus, the dataset is suitable for implementing structured imputation methods, especially those that leverage temporal patterns and daily seasonal variations present in environmental data.

After evaluating the dataset's completeness and trustworthiness, it is crucial to analyze the temporal properties of the recorded variables to guarantee the consistency and interpretability of the observed patterns. Figure 2 illustrates the temporal fluctuations of ambient temperature recorded from 08 May 2023 to 08 June 2023. Despite the y-axis range in the original plot extending to 700 °C, the recorded temperatures remained between 0 and 40 °C, indicating that the sensor device functioned under standard climatic conditions. The confined and consistent range verifies that no atypical thermal incidents transpired during the measurement interval (Laha et al., 2022). The lack of sudden spikes or adverse deviations commonly linked to sensor malfunctions further indicates that the data acquisition module operated successfully during the monitoring period (Laha et al., 2022; Mois et al., 2017).

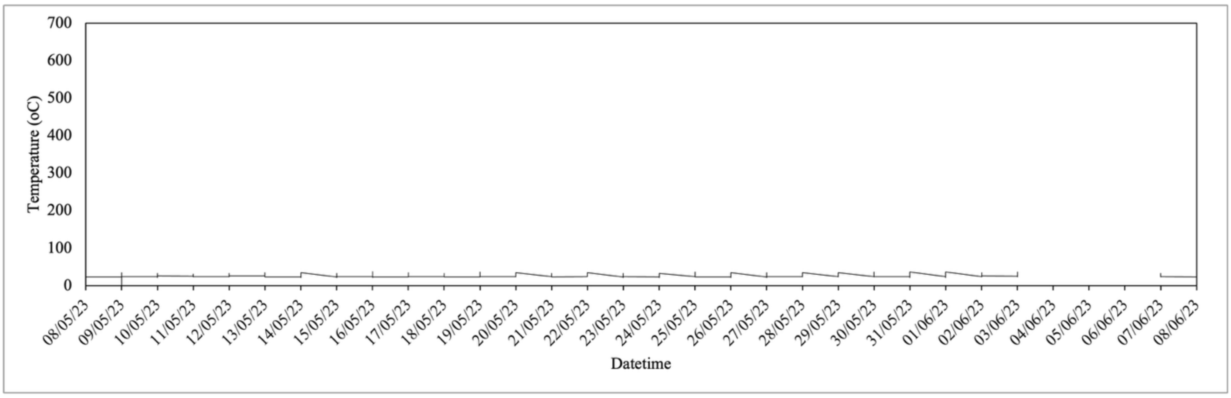

Figure 2. Raw Temperature Dataset

A distinct diurnal trend is evident across the sample. Temperature values consistently rise during daylight and decline post-sunset, establishing a regular daily cycle. The early period often exhibits the lowest temperatures, which progressively rise to a noon zenith before declining again throughout the evening and into the midnight hours. This clearly delineated circadian thermal behavior aligns with standard outdoor ambient conditions, indicating that the dataset possesses a robust, consistent seasonal component. This stability is advantageous for later analytical tasks such as seasonal decomposition, forecasting, and pattern-based imputation.

This consistency suggests that time-of-day-based imputation methods, such as categorization by hour and minute, are both conceptually and empirically consistent with the data's inherent structure (Fatyanosa et al., 2023; Seu et al., 2022). The robustness of the daily cycle reinforces the statistical premise that temperatures at the same time across different dates exhibit similar distributions. These attributes jointly affirm that the dataset is of superior quality and well-suited for sophisticated time-series analytics, encompassing imputation, anomaly detection, and predictive modelling (Dau et al., 2018; Yu et al., 2014).

Notwithstanding the considerable intra-day fluctuations, the inter-day average humidity is consistent, typically sustaining levels between 75–85% RH throughout the month. This consistency indicates that the regional atmospheric moisture conditions during the observed period were climatologically stable, devoid of significant weather anomalies that would cause abrupt changes in mean humidity levels. The lack of abrupt discontinuities or artificial spikes suggests that the sensing system functioned successfully, with no indications of signal dropouts, ADC saturation, or digital communication failures, as are typically associated with low-cost humidity sensors (Hamel et al., 2024; Vaicdan et al., 2019).

Several days show significant declines in humidity, represented by elongated downward vertical lines in the graph. These vertical fluctuations correspond to early-afternoon aridity episodes typically induced by heightened thermal activity and air mixing. This phenomenon is prevalent in tropical and subtropical climes, characterized by elevated early-morning humidity that subsequently diminishes sharply under intense sunlight (Budiwati et al., 2022; Indrawati et al., 2024). These patterns further confirm that the dataset reflects genuine environmental variability rather than measurement noise. The regular periodicity indicates that time-of-day analytical methods, such as seasonal imputation, hourly aggregation, or harmonic decomposition, are suitable and theoretically substantiated by the data's inherent structure (Indrawati et al., 2024). The robust daily consistency also ensures the efficient use of statistical models that depend on temporal autocorrelation.

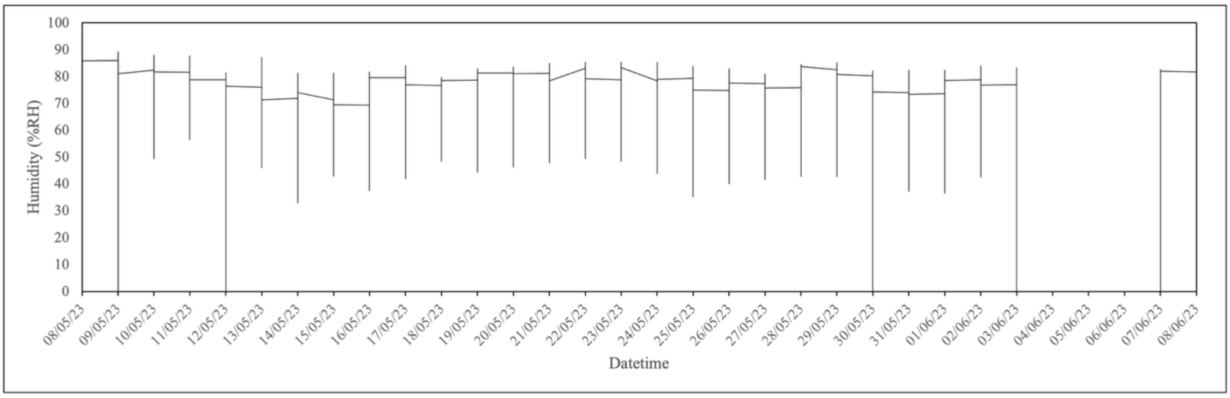

Figure 3. Raw Humidity Dataset

To better understand the temporal trends in temperature and humidity measurements, it is essential to analyze the dataset's statistical properties before imputation. The descriptive statistics presented in Table 2 provide further insight into the central tendency, dispersion, and distributional characteristics of the raw sensor data, thereby facilitating data quality assessment and informing the selection of appropriate preprocessing and imputation techniques.

Table 2. IoT Characteristic Data Pre-encoded

Parameters	count	mean	std	min	25%	50%	75%	max
Temp	120,740	26.52	4.58	00.00	23.03	25.03	29.06	626.66
Hum	120,738	68.68	12.86	00.00	58.01	72.09	79.00	89.40

The interquartile range (23.03–29.06°C) indicates that most temperature measurements are concentrated within a small, climatologically valid range. The recorded low of 0.00°C and maximum of

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

626.66°C indicate potential sensor malfunctions or transmission problems, as these values are physically improbable in the monitoring scenario (Nassif et al., 2021). These abnormalities necessitate comprehensive preprocessing and imputation procedures to correct distorted or erroneous measurements.

The humidity dataset has 120,738 observations, with a mean of 68.68% RH and a standard deviation of 12.86% RH, indicating greater intrinsic volatility in relative humidity than in temperature. The interquartile range (58.01–79.00% RH) indicates a predominantly stable distribution; however, the minimum value of 0.00% RH is a distinct outlier, likely due to sensor malfunction or communication failure. In contrast to temperature, the humidity maxima remain within a plausible environmental range (89.40% RH), further substantiating the conclusion that the dataset, although predominantly reliable, includes isolated erroneous values that require rectification.

The statistics in Table 2 collectively indicate that the dataset is generally typical of actual environmental conditions; however, it includes some abnormal values that require meticulous preprocessing. This underscores the necessity of employing structured imputation methods that leverage robust diurnal patterns already established, ensuring that subsequent analyses and modeling endeavors rely on high-quality, physically relevant data.

Figure 4 illustrates the refined temperature time-series from 08 May 2023 to 08 June 2023, after the removal of extreme anomalies identified in Table 2. The resulting pattern demonstrates a far more coherent and environmentally reasonable trajectory, with temperature values consistently fluctuating between approximately 20°C and 35°C, corresponding with normal outdoor ambient circumstances during the measurement period. The elimination of outliers, namely the improbable maximum result of 626.66°C, has successfully reinstated the anticipated thermal profile and eradicated aberrations that could skew later assessments (Bhuyan et al., 2014; Nassif et al., 2021).

The visual consistency observed over days indicates significant temporal autocorrelation and seasonal regularity, which are essential for the reliability of time-of-day imputation techniques and the efficacy of time-series forecasting models. The synchronization of peak timing and amplitude across several dates substantiates the premise that temperature at a specific hour is statistically analogous across multiple days—a fundamental prerequisite for the hour–minute grouping technique utilized in this work (Mahajan et al., 2020; Salvatore et al., 2002).

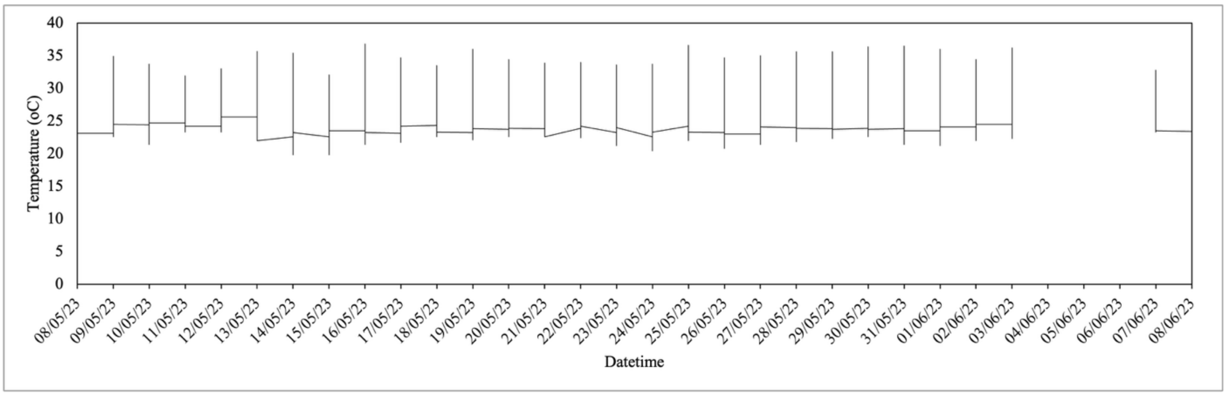

Figure 4. The improved temperature time series dataset.

In accordance with the temperature dataset, it is imperative to assess the behavior of the humidity measurements after eliminating anomalous and corrupted values. This evaluation offers a more precise depiction of the atmospheric moisture dynamics recorded during the observation period. Figure 6 presents the improved humidity time series, facilitating a more precise analysis of its diurnal fluctuation and overall quality.

Figure 5 displays the refined humidity readings the period of measurement, demonstrating a significantly enhanced, more consistent profile compared to the unprocessed dataset. Upon eliminating implausible minimum values (0% RH) and stabilizing discontinuities, the data reveal humidity levels primarily within the 65–

85% RH range, consistent with the region's actual environmental conditions throughout the observation period. The preprocessing phase has successfully reduced sensor dropouts and communication faults, yielding a more coherent and interpretable signal .

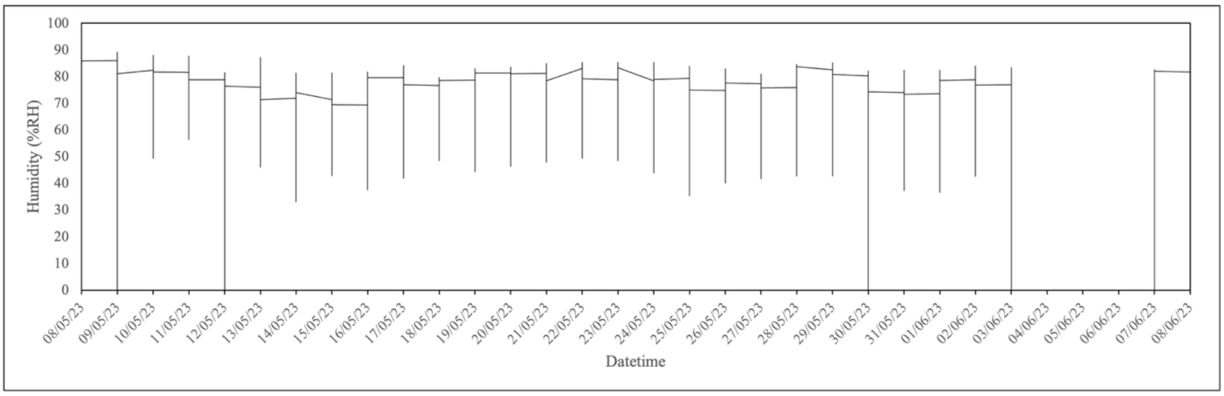

Figure 5. The improved humidity time series

The revised series continues to exhibit significant intra-day fluctuation, as indicated by the vertical ranges denoting daily minimum and maximum humidity levels. The variations—typically between 40% and 90% RH during a single day—align with established atmospheric dynamics influenced by temperature changes, solar radiation, nocturnal cooling, and dew formation. The elimination of inaccurate zero-humidity values has eradicated artificial troughs that previously skewed the dataset, allowing the natural diurnal moisture cycle to manifest more distinctly.

Although more volatile than temperature, the humidity profile has a stable inter-day pattern, with daily peaks and troughs occurring within comparable amplitude ranges. This consistency indicates robust diurnal seasonality, confirming that humidity follows a discernible 24-hour cycle. This temporal structure substantiates the theoretical rationale for using hour- and minute-based seasonal imputation techniques, as humidity data at specific times of day exhibit analogous patterns throughout the monitoring period .

A statistical overview of the dataset following outlier removal is provided in Table 3 to augment the visual evaluation of the cleaned temperature and humidity time series. This summary offers a quantitative analysis of how the -

Table 3. Characteristic Data of IoT Post-Cleaning

Parameters	count	mean	std	min	25%	50%	75%	max
Temp	120,671	26.53	3.83	26.53	23.30	25.30	29.60	36.90
Hum	120,673	68.71	12.77	68.71	58.10	73.00	79.00	89.40

-preparation processes enhanced data quality by removing physically impossible or inconsistent findings.

Table 3 shows the dataset after removing 69 temperature and 65 humidity outliers, confirming that the cleaning process successfully restored physical coherence and reduced noise. The temperature variable now contains 120,671 valid entries with a mean of 26.53°C, a lower standard deviation of 3.83°C, and plausible minimum–maximum values (23.30–36.90°C), indicating that the excluded anomalies had disproportionately inflated variability. The humidity data, comprising 120,673 valid observations, shows a mean of 68.71% RH and a standard deviation of 12.77% RH, with an interquartile range of 58.10–79.00% RH and a maximum of 89.40% RH, reflecting realistic environmental conditions after removing mostly implausible low readings caused by sensor dropout. With the dataset cleaned and statistically consistent, the next step is to evaluate whether the imputation method preserves temporal continuity in the reconstructed temperature series, ensuring alignment with natural diurnal patterns and avoiding the introduction of artificial behaviors.

Figure 6 illustrates the temperature time series after applying the seasonal time-of-day imputation method, showing a smooth trajectory that closely follows the natural diurnal cycle present in the cleaned dataset. The reconstructed

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

values remain within the expected ambient temperature range of 22–33°C, and the imputed segments—particularly visible in the rightmost portion of the figure—blend seamlessly with surrounding observations, preserving both local continuity and the broader temporal pattern. The graphic also confirms that the method accurately reproduces daily minima at night and maxima during daytime, with no abrupt jumps or artificial plateaus, demonstrating the effectiveness of the hour–minute grouping approach in leveraging strong diurnal seasonality, consistent with findings such as (Kwok et al., 2023). Unlike raw gaps or oversmoothed flatline artifacts produced by simpler imputation techniques, the reconstructed series retains natural variability, including subtle fluctuations and shifts in peak timing characteristic of real environmental dynamics. This preservation of physical realism ensures that the imputed temperature data remain suitable for downstream tasks such as anomaly detection, forecasting, and sensor reliability assessment (Nassif et al., 2021)

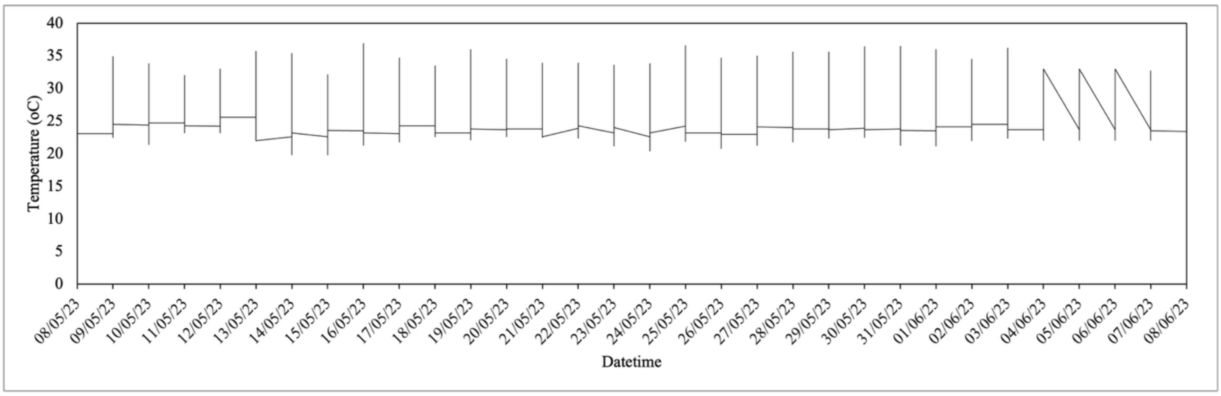

Figure 6. Temperature time series after implementing the seasonal imputer.

After evaluating the imputed temperature series, it is also crucial to assess the imputation method's efficacy on the humidity observations, which are inherently more unpredictable and susceptible to environmental changes. This phase guarantees that the imputed values uphold physical plausibility and align with the diurnal patterns evident in the cleaned dataset. Figure 8 displays the reconstructed humidity time series, facilitating a thorough evaluation of the method's ability to recover missing parts.

Figure 7 illustrates the humidity time series after implementing the seasonal, time-of-day-based imputation technique. The imputed series demonstrates a continuous, environmentally realistic trajectory, sustaining humidity values primarily between 65% RH and 85% RH—aligned with the natural atmospheric conditions observed in the cleaned data. The technique effectively prevented the reintroduction of unrealistic values, such as the improbable 0% RH readings found in the raw dataset, demonstrating that the imputation process maintained physical realism.

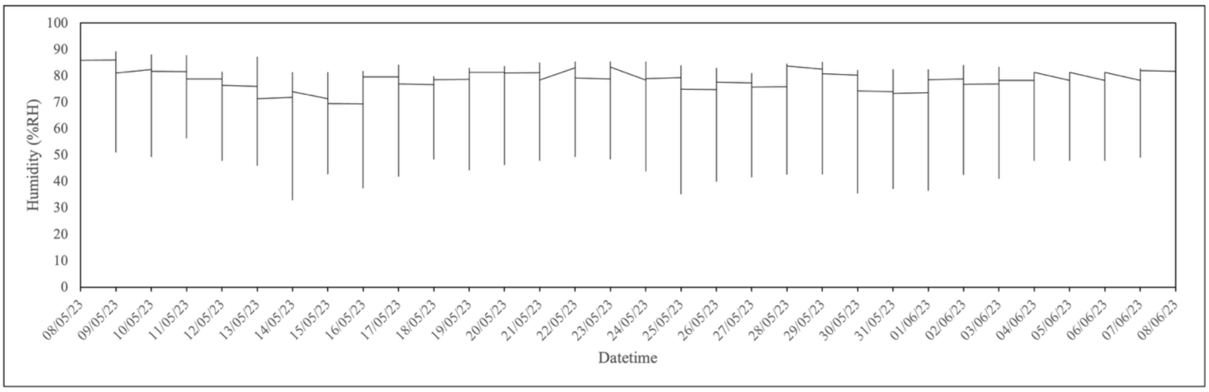

Figure 7. Humidity time series after implementing the seasonal imputer.

The rebuilt humidity series preserves the expected diurnal structure despite its inherent intra-day volatility, with early-afternoon minima and overnight maxima clearly maintained, and the imputed segments blending smoothly into surrounding observations without abrupt transitions or artificial patterns—demonstrating the strength of the hour–minute grouping approach in exploiting the data's periodicity.

The stable daily amplitude ranges further indicate that the method retains natural variability instead of oversmoothing the signal, an essential property for downstream tasks such as anomaly detection, classification, and forecasting. Completeness checks before and after reconstruction also confirm that the imputation method successfully resolved all missing values without introducing new gaps or inconsistencies, yielding a coherent and fully restored humidity dataset.

Table 4 demonstrates that the imputation process successfully eliminated all missing values in both the temperature and humidity data, reducing the initial 15,955 missing temperature records and 15,953 missing humidity records—typical of long-term IoT deployments affected by connectivity or sensor issues—to zero. This complete restoration of missing segments ensures full temporal continuity, which is essential for downstream tasks such as anomaly detection, feature extraction, and predictive modeling that depend on uninterrupted time-series inputs.

Table 4. Absence of post-imputed counts

Parameters	Missing counts before imputation	Imputation
Temperature	15,955	0
Humidity	15,953	0

The ability to recover all gaps without introducing new inconsistencies further indicates that the hour–minute seasonal imputation method effectively captured the dataset's inherent periodic structure, producing values that remain statistically and physically coherent. With the missing data fully resolved, the next step involves evaluating how the imputation influences the dataset's statistical properties, ensuring that the reconstructed values remain consistent with the natural behavior of the monitored environmental variables. Table 5 provides a summary of the descriptive statistics of the fully imputed dataset to support this assessment.

Table 5. Characteristic data following imputation.

Parameters	count	mean	std	min	25%	50%	75%	max
Temp	136,626	26.52	3.81	19.8	23.3	25.2	29.7	36.90
Hum	136,626	68.74	12.63	33.0	58.1	73.3	78.9	89.40

Table 5 presents the statistical attributes of the temperature and humidity variables after imputing all missing values, yielding a comprehensive dataset of 136,626 records for each parameter. The mean temperature is 26.52°C, almost identical to the pre-imputation mean, indicating that the reconstruction procedure maintained the data's central tendency. The standard deviation marginally decreased to 3.81°C, indicating reduced variability while preserving realistic diurnal swings.

The dataset reveals a mean humidity of 68.74% RH and a standard deviation of 12.63% RH, both of which closely align with the distribution prior to imputation. The retained interquartile range (IQR: 58.1–78.9% RH) indicates that the inherent variability of humidity was predominantly preserved. The low value of 33% RH and the maximum of 89.40% RH suggest that the imputation method did not yield extreme values or overly smoothed segments. The results indicate that the seasonal, time-of-day imputation method successfully addressed the missing entries without compromising the overall statistical integrity of the data.

A controlled missing-rate experiment was conducted to assess the robustness of the proposed imputation method, during which variable amounts of temperature data were routinely omitted and later reconstructed. This process facilitates an objective evaluation of the method's stability under varying degrees of data incompleteness. The consolidated findings of this experiment, encompassing both baseline errors in the unprocessed masked data and post-imputation errors, are displayed in Table 6.

Table 6 illustrates the efficacy of the proposed imputation method over five scenarios of

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

missing rates, varying from 1% to 50%. The baseline reconstruction errors derived from the masked raw data persistently exhibit elevated levels, with MAE values ranging from 3.14 to 3.15°C, RMSE values between 4.03 and 4.05°C, and MAPE values of about 11.1 to 11.2%, indicating that the dataset is significantly compromised when data are omitted without imputation.

The imputed findings exhibit significantly enhanced accuracy at all levels of missingness. The imputation approach attains MAE values of roughly 0.86–0.87°C, RMSE values of 1.36–1.37°C, and MAPE values of 3.17–3.22%, indicating a three- to fourfold improvement over the baseline. Notably, these performance indicators exhibit significant stability even when the missing rate reaches 50%, with relatively little fluctuation across repeated trials.

Table 6. Temperature omission rate test results

Missing rate	MAE Raw		RMSE Raw		MAPE Raw		MAE Imputed		RMSE Impute		MAPE Imputed	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
1%	3.15	0.08	4.05	0.09	11.16	0.24	0.87	0.02	1.37	0.04	3.18	0.08
5%	3.14	0.03	4.03	0.04	11.11	0.09	0.86	0.01	1.36	0.02	3.17	0.04
10%	3.14	0.02	4.04	0.03	11.11	0.07	0.86	0.01	1.36	0.01	3.18	0.02
20%	3.14	0.01	4.03	0.02	11.11	0.04	0.87	0.01	1.36	0.01	3.19	0.02
50%	3.14	0.00	4.03	0.01	11.11	0.01	0.87	0.00	1.36	0.00	3.22	0.01

The exceptionally low standard deviations underscore the method's durability and reproducibility, affirming that the hour–minute seasonal grouping strategy efficiently leverages the pronounced diurnal structure of temperature data.

Table 7 illustrates the reconstruction accuracy of the proposed imputation approach for the humidity variable across missing rates ranging from 1% to 50%. The baseline errors of the masked raw data persistently exhibit elevated levels, with MAE values ranging from 10.41 to 10.47%RH, RMSE values between 13.36 and 13.43%RH, and MAPE values between 18.18 and 18.32%, indicating that the exclusion of observations without imputation significantly distorts the data.

Table 7. Results of the humidity missing rate test

Missing Rate	MAE Raw		RMSE Raw		MAPE Raw		MAE Imputed		RMSE Imputed		MAPE Imputed	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
1%	10.41	0.18	13.36	0.19	18.18	0.39	3.92	0.11	5.73	0.12	6.22	0.19
5%	10.44	0.11	13.39	0.13	18.25	0.25	3.94	0.05	5.76	0.06	6.26	0.09
10%	10.45	0.10	13.41	0.11	18.28	0.24	3.95	0.03	5.75	0.03	6.28	0.05
20%	10.47	0.07	13.43	0.08	18.32	0.16	3.97	0.02	5.78	0.02	6.32	0.04
50%	10.46	0.02	13.42	0.03	18.32	0.05	4.01	0.01	5.80	0.02	6.39	0.02

The imputed findings indicate a notable improvement in accuracy, with MAE decreasing to roughly 3.92–4.01% RH, RMSE decreasing to 5.73–5.80% RH, and MAPE decreasing to 6.22–6.39%. Despite these post-imputation errors being larger than those observed in the temperature experiments—indicative of the naturally higher volatility of humidity—the enhancement relative to the baseline is significant across all levels of missingness. The durability of the imputation findings is demonstrated by the negligible standard deviations and limited fluctuations in error metrics, even at a missing rate of 50%. The findings indicate that the seasonal hour-minute-based imputation method effectively leverages the pronounced diurnal pattern in humidity data, enabling the reliable restoration of missing segments despite significant intra-day fluctuations. The results affirm that the method is accurate and resilient for humidity imputation, demonstrating high performance across different levels of data incompleteness.

Table 8 presents the 95% confidence intervals (CI) of the imputation error metrics for temperature across different missing rates. The results indicate that the estimated errors remain highly stable under varying levels of missing data. Specifically, the MAE values range from 10.41 to 10.47, while the RMSE values vary only slightly between 13.36 and 13.43, and the MAPE values remain within 18.18–18.32. This narrow variation suggests that the imputation process produces consistent reconstruction performance regardless of the missing data proportion.

Table 8. Confidence Intervals of Temperature Error Metrics Across Different Missing Rates

Missing Rate	MAE Imputed (95% CI)	RMSE Imputed (95% CI)	MAPE Imputed (95% CI)
1%	10,41 ± 0.0069	13,36 ± 0.0076	18,18 ± 0.0151
5%	10,44 ± 0.0043	13,44 ± 0.0049	18,25 ± 0.0099
10%	10,45 ± 0.0040	13,41 ± 0.0045	18,28 ± 0.0092
20%	10,47 ± 0.0027	13,43 ± 0.0031	18,32 ± 0.0062
50%	10,46 ± 0.0009	13,42 ± 0.0010	18,32 ± 0.0020

Furthermore, the confidence intervals are very small, with MAE margins decreasing from ±0.0069 at 1% missing rate to ±0.0009 at 50% missing rate. A similar pattern is observed for RMSE and MAPE, where the CI ranges gradually become narrower as the missing rate increases. This indicates that the variability of the experimental results is minimal and that the imputation method maintains high statistical stability across repeated simulations.

Table 9 confidence intervals (CI) of the imputation error metrics for humidity across different missing rates. The results indicate that the reconstruction errors remain relatively stable as the proportion of missing data increases. The MAE values vary slightly from 10.41 to 10.47, while the RMSE values range between 13.36 and 13.44, and the MAPE values remain within 18.18–18.32. This limited variation suggests that the imputation approach consistently maintains similar reconstruction accuracy across different missing-rate scenarios.

Table 9. Confidence Intervals of Humidity Error Metrics Across Different Missing Rates

Missing Rate	MAE Imputed (95% CI)	RMSE Imputed (95% CI)	MAPE Imputed (95% CI)
1%	10,41 ± 0.0069	13,36 ± 0.0076	18,18 ± 0.0151
5%	10,44 ± 0.0043	13,44 ± 0.0049	18,25 ± 0.0099
10%	10,45 ± 0.0040	13,41 ± 0.0045	18,28 ± 0.0092
20%	10,47 ± 0.0027	13,43 ± 0.0031	18,32 ± 0.0062
50%	10,46 ± 0.0009	13,42 ± 0.0010	18,32 ± 0.0020

In addition, the confidence intervals are notably narrow, with the MAE margin decreasing from ±0.0069 at a 1% missing rate to ±0.0009 at a 50% missing rate. A comparable trend is observed for RMSE and MAPE, where the CI ranges become progressively smaller as the missing rate increases. This pattern indicates that the variability across repeated simulations is minimal, demonstrating the statistical stability of the imputation results.

Compared with (Fatyanosa et al., 2023), the dataset used in this study differs, and the anomaly imputation in that work was not conducted within a real-world IoT-based environmental monitoring scenario. Meanwhile, (Juna et al., 2022) focused primarily on computational modeling for water quality prediction using a multilayer perceptron, whereas the present study evaluates the performance of imputation methods under varying missing-rate scenarios using seasonal input patterns. In comparison with (Kwok et al., 2023; Qin et al., 2021), although similar parameters and evaluation approaches are applied, this research focuses on environmental monitoring related to acid rain using atmospheric parameters obtained directly from an IoT-based sensing system. Therefore, the evaluation results obtained in this study demonstrate promising performance within the context of real IoT environmental monitoring applications.

Overall, the narrow confidence intervals across all missing-rate scenarios of this study indicate that the proposed imputation method maintains stable and reliable reconstruction performance for both temperature and humidity data. This confirms the robustness of the approach in handling missing values in environmental IoT sensor time series.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study demonstrates that the proposed preprocessing pipeline, which combines anomaly cleaning and seasonal time-of-day imputation, effectively improves the quality and continuity of environmental IoT sensor data. The removal of extreme outliers followed by seasonal imputation successfully reconstructs missing observations while preserving the natural diurnal dynamics of temperature and humidity time series.

Experimental results show that the proposed approach achieves consistently low reconstruction errors across different missing-rate scenarios. For temperature data, the method maintains MAE values between 0.86–0.87°C and MAPE between 3.17–3.22%, while humidity errors are significantly reduced from approximately 10.45%RH to 3.92–4.01%RH. The narrow 95% confidence intervals further indicate that the reconstruction performance remains stable and statistically reliable across repeated simulations, even under high missingness conditions.

Overall, the findings confirm that the seasonal imputation approach provides a robust and computationally efficient solution for handling missing values in environmental IoT sensor time series, enabling the generation of reliable datasets for environmental monitoring and further analytical applications.

Suggestion

Future work may focus on enhancing the proposed seasonal imputation framework by integrating multi-sensor information, adopting adaptive or learning-based seasonal modeling to accommodate environmental variability, and validating the method across diverse climatic regions to improve generalizability. Additionally, exploring hybrid approaches that combine seasonal patterns with machine-learning models—such as LSTM autoencoders or probabilistic time-series methods—could further strengthen reconstruction accuracy, especially under high missingness. Finally, developing lightweight, real-time implementations suitable for edge or embedded IoT devices would enable on-device data correction and improve the practicality of environmental monitoring systems.

REFERENCES

Bhattacharya, S., Kamper, F., & Beirlant, J. (2023). Outlier detection based on extreme value theory and applications. Scandinavian Journal of Statistics, 50(3), 1466–1502. https://doi.org/10.1111/sjos.12665

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2014). Network anomaly detection: Methods, systems and tools. IEEE Communications Surveys and Tutorials, 16(1), 303–336. https://doi.org/10.1109/SURV.2013.052213.00046

Boukerche, A., Zheng, L., & Alfandi, O. (2020). Outlier Detection: Methods, Models, and Classification. ACM Computing Surveys, 53(3). https://doi.org/10.1145/3381028

Budiwati, T., Tanti, D. A., Indrawati, A., & Santoso, M. (2022). Analisis Kualitas Udara dan Hubunganya dengan Deposisi Basah di Semarang. In Dinamika dan Teknologi Atmosfer Benua Maritim Indonesia (pp. 185–200).

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Dau, H. A., Silva, D. F., Petitjean, F., Forestier, G., Bagnall, A., Mueen, A., & Keogh, E. (2018). Optimizing dynamic time warping's window width for time series data mining applications. Data Mining and Knowledge Discovery, 32(4), 1074–1120. https://doi.org/10.1007/s10618-018-0565-y

Eckardt, M., & Kerber, W. (2024). Property rights theory, bundles of rights on IoT data, and the EU Data Act. European Journal of Law and Economics. https://doi.org/10.1007/s10657-023-09791-8

Fatyanosa, T. N., Data, M., Firdausanti, N. A., Prayoga, P. H. N., Mendonça, I., & Aritsugi, M. (2023). ImputAnom: Anomaly Detection Framework Using Imputation Methods for Univariate Time Series. In P. Delir Haghighi, E. Pardede, G. Dobbie, V. Yogarajan, N. A. S. ER, G. Kotsis, & I. Khalil (Eds.), Information Integration and Web Intelligence (pp. 56–61). Springer Nature Switzerland.

Fatyanosa, T. N., Firdausanti, N. A., Prayoga, P. H. N., Kuriu, M., Aritsugi, M., & Mendonça, I. (2024). Meta-learning for vessel time series data imputation method recommendation. Expert Systems with Applications, 251. https://doi.org/10.1016/j.eswa.2024.124016

Hamel, P., Ding, N., Cherqui, F., Zhu, Q., Walcker, N., Bertrand-Krajewski, J. L., Champrasert, P., Fletcher, T. D., McCarthy, D. T., Navratil, O., & Shi, B. (2024). Low-cost monitoring systems for urban water management: Lessons from the field. Water Research X, 22. https://doi.org/10.1016/j.wroa.2024.100212

Hudda, S., & Haribabu, K. (2025). A review on WSN based resource constrained smart IoT systems. Discover Internet of Things, 5(1), 56. https://doi.org/10.1007/s43926-025-00152-2

Indrawati, A., Tanti, D. A., Ambarsari, N., Ma'ruf, I. F., Sumaryati, Setyawati, W., Pusfitasari, E. D., Nugroho, G. A., Cholianawati, N., Sinatra, T., Halimurrahman, Triani, L. R., Nafiisyanti, A., Putra, S. I., Nurlatifah, A., Huo, M., & Awaludin, A. (2024). Spatiotemporal distribution in chemical composition of wet atmospheric deposition in Bandung Indonesia. Environmental Science and Pollution Research. https://doi.org/10.1007/s11356-024-35485-y

Jalasena, A., Rachmawati, L. M., Hadi Syahputra, M. F., Hasmul, N. A., Suwandi, Indah Utami, A. R., Nugroho, B. S., & Chandra, I. (2023). PRELIMINARY STUDY OF URGENCY TO MONITOR INDOOR AIR QUALITY AT TELKOM EDUCATION AREA IN THE POST-PANDEMIC. IOP Conference Series: Earth and Environmental Science, 1194(1). https://doi.org/10.1088/1755-1315/1194/1/012004

Kwok, C.-F., Qian, G., & Kuleshov, Y. (2023). Analyzing Error Bounds for Seasonal-Trend Decomposition of Antarctica Temperature Time Series Involving Missing Data. Atmosphere, 14(2). https://doi.org/10.3390/atmos14020193

Laha, S. R., Pattanayak, B. K., & Pattnaik, S. (2022). Advancement of Environmental Monitoring System Using IoT and Sensor: A Comprehensive Analysis. AIMS Environmental Science, 9(6), 771–800. https://doi.org/10.3934/environsci.2022044

Lee, J. H., & Charles Huber, J. (2021). Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? In Iran J Public Health (Vol. 50, Issue 7). https://creativecommons.org/licenses/by-nc/4.0/

Mahajan, M., Kumar, S., Pant, B., & Tiwari, U. K. (2020). Incremental Outlier Detection in Air Quality Data Using Statistical Methods. 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 1–5. https://doi.org/10.1109/ICDABI51230.2020.9325683

Mara Ribeiro, S. I., & Leite de Castro, C. I. (2022). Missing Data in Time Series: A Review of Imputation Methods and Case Study. In Journal of the Brazilian Society on Computational Intelligence (SBIC) (Vol. 20, Issue 1).

Mohammed, M. B., Zulkafli, H. S., Adam, M. B., Ali, N., & Baba, I. A. (2021). Comparison of five imputation methods in handling missing data in a continuous frequency table. AIP Conference Proceedings, 2355(1), 040006. https://doi.org/10.1063/5.0053286

Mois, G., Folea, S., & Sanislav, T. (2017). Analysis of Three IoT-Based Wireless Sensors for Environmental Monitoring. IEEE Transactions on Instrumentation and Measurement, 66(8), 2056–2064. https://doi.org/10.1109/TIM.2017.2677619

Muhr, D., Affenzeller, M., & Küng, J. (2023). A Probabilistic Transformation of Distance-Based Outliers. Machine Learning and Knowledge Extraction, 5(3), 782–802. https://doi.org/10.3390/make5030042

Mustofa, A. H., Chandra, I., Salam, R. A., & Rachmawati, L. M. (2022). Vertical Structure of PM2.5and CO2Concentrations in the Greater Bandung Air Basin. Journal of Physics: Conference Series, 2377(1). https://doi.org/10.1088/1742-6596/2377/1/012052

Nassif, A. B., Talib, M. A., Nasir, Q., & Dakalbab, F. M. (2021). Machine Learning for Anomaly Detection: A Systematic Review. In IEEE Access (Vol. 9, pp. 78658–78700). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2021.3083060

Pham, T. M., Pandis, N., & White, I. R. (2022). Missing data, part 2. Missing data mechanisms:

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i2.475
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Missing completely at random, missing at random, missing not at random, and why they matter. American Journal of Orthodontics and Dentofacial Orthopedics, 162(1), 138–139. https://doi.org/10.1016/j.ajodo.2022.04.001

Qin, Y., Ren, G., Zhang, P., Wu, L., & Wen, K. (2021). An imputation method for the climatic data with strong seasonality and spatial correlation. Theoretical and Applied Climatology, 144(1–2), 203–213. https://doi.org/10.1007/s00704-021-03537-9

Ramadhan, A., Chandra, I., Indrawati, A., Tanti, D. A., Mendonça, I., & Aritsugi, M. (2025). Decision support system using the expert-fuzzy method in validating real-time acid rain measurements. Journal of Reliable Intelligent Environments, 11(4). https://doi.org/10.1007/s40860-025-00258-y

Ramadhan, A., Indrawati, A., Tanti, D. A., Setyawati, W., & Chandra, I. (2025). AI-Driven Imputation and Correlation Analysis of Environmental Parameters for Understanding pH Variability in Acid Rain. 2025 IEEE International Conference on Networking, Intelligent Systems, and IoT (ICONS-IoT), 75–81. https://doi.org/10.1109/ICONS-IoT65216.2025.11211132

Saleem, S., Aslam, M., & Rukh Shaukat, M. (2021). A REVIEW AND EMPIRICAL COMPARISON OF UNIVARIATE OUTLIER DETECTION METHODS (Vol. 37, Issue 4).

Salvatore, D., Reagle, D., & New York Chicago San Francisco Lisbon London Madrid Mexico City Milan New Delhi San Juan Seoul Singapore Sydney Toronto, M.-H. (2002). Theory and Problems of STATISTICS AND ECONOMETRICS SECOND EDITION Schaum's Outline Series. https://doi.org/10.1036/0071395687

Sarker, I. H. (2021). Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. In SN Computer Science (Vol. 2, Issue 5). Springer. https://doi.org/10.1007/s42979-021-00765-8

Seo, S. (2006). A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets. University of Pittsburgh.

Seu, K., Kang, M.-S., & Lee, H. (2022). An Intelligent Missing Data Imputation Techniques: A Review. www.joiv.org/index.php/joiv

Shadbahr, T., Roberts, M., Stanczuk, J., Gilbey, J., Teare, P., Dittmer, S., Thorpe, M., Torné, R. V., Sala, E., Lió, P., Patel, M., Preller, J., Lió, P., Walton, N., Holzer, M., Wassin, M., Escudero Sánchez, L., Babar, J., Prosch, H., … Schönlieb, C. B. (2023). The impact of imputation quality on machine learning classifiers for datasets with missing values. Communications Medicine, 3(1). https://doi.org/10.1038/s43856-023-00356-z

Singh, D., Dahiya, M., Kumar, R., & Nanda, C. (2021). Sensors and systems for air quality assessment monitoring and management: A review. Journal of Environmental Management, 289, 112510. https://doi.org/https://doi.org/10.1016/j.jenvman.2021.112510

Smiti, A. (2020). A critical overview of outlier detection methods. Computer Science Review, 38, 100306. https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100306

Vaicdan, F., Chandra, I., & Suhendi, A. (2019). Pengamatan Konsentrasi Massa PM2.5 di Cekungan Udara Bandung Raya (Observation of PM2.5 mass concentrations in the Greater Bandung air basin).

Vasenin, D., Pasetti, M., Astolfi, D., Savvin, N., Rinaldi, S., & Berizzi, A. (2024). Incorporating Seasonal Features in Data Imputation Methods for Power Demand Time Series. IEEE Access, 12, 103520–103536. https://doi.org/10.1109/ACCESS.2024.3434652

Yu, Y., Zhu, Y., Li, S., & Wan, D. (2014). Time series outlier detection based on sliding window prediction. Mathematical Problems in Engineering, 2014. https://doi.org/10.1155/2014/879736

Zhao, Y., & Udell, M. (2024). gcimpute: A Package for Missing Data Imputation. Journal of Statistical Software, 108(4), 1–27. https://doi.org/10.18637/jss.v108.i04