Journal of E-Technology

DLINE JOURNALS

Print ISSN: 0976-3503
Online ISSN: 0976-2930

JET 2026: 17 (3)
https://doi.org/10.6025/jet/2026/17/3/103-121

Multiscale Time-Series Analysis of Temperature Variability in Ho Chi Minh City: Spectral, Seasonal, and Wavelet-Based Characterization

Nguyen Minh Tuan
Posts and Telecommunications Institute of Technology (PTIT)
11 Nguyen Dinh Chieu, Sai Gon Ward, Ho Chi Minh City
Vietnam
minh.tuan@ptithcm.edu.vn

ABSTRACT

This study investigates the multiscale temporal dynamics of near surface temperature variability in Ho Chi Minh City, Vietnam, utilizing a comprehensive three-year hourly meteorological dataset comprising 28,508 observations. A unified analytical framework integrating time domain, frequency domain, and time frequency techniques was applied to robustly isolate periodicities and assess structural stability. Methodologies included Fast Fourier Transform (FFT), Fisher's g-test, Monte Carlo surrogate modeling, stationarity diagnostics (ADF and KPSS tests), autocorrelation analysis, classical seasonal decomposition, and Continuous Wavelet Transform (CWT) coupled with multi resolution decomposition.

Results reveal a highly significant, deterministic 24-hour diurnal cycle governing temperature, relative humidity, and precipitation, primarily driven by intense solar forcing and regional land sea breeze regimes. Conversely, mean sea level pressure is dominated by low frequency seasonal fluctuations of approximately 1.08 years. Stationarity tests confirm the dataset's inherent non stationarity, while autocorrelation diagnostics reveal strong temporal persistence and long memory. Furthermore, wavelet scalograms and multi-resolution decomposition illustrate that while long term climatic forcing accounts for the overwhelming majority of statistical variance, short term weather anomalies introduce critical high frequency fluctuations.

These findings underscore the limitations of standalone linear models for environmental forecasting. The identified hierarchical temporal structure strongly supports the development of hybrid predictive architectures. Future forecasting systems should integrate wavelet based multi resolution feature extraction, routing long term trends through deep learning frameworks while optimizing short term residuals via machine learning ensembles to significantly enhance urban climate and energy demand predictions.

Keywords: Multiscale Time-Series Analysis, Temperature Variability, Spectral Analysis, Continuous Wavelet Transform (CWT), Multi-resolution Decomposition, Non-Stationarity, Diurnal Cycle, Tropical Monsoon Climate

Received: 19 September 2025, Revised 9 January 2026, Accepted 4 March 2026

1. Introduction

The Earth's climate system is highly dynamic and interconnected, comprising five major subsystems: the atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere [1, 2]. Processes occurring within and among these subsystems operate across a wide range of spatial and temporal scales, extending from molecular interactions to planetary scale phenomena and from seconds to millions of years [3]. These components continuously exchange mass, energy, and momentum, creating complex feedback mechanisms and cascade effects that influence the evolution of climate states. Although the climate system is generally regarded as a closed system with respect to mass exchange, it remains open to energy transfers, thereby allowing external influences to affect its behaviour [4].

2. Earlier Studies

Climate dynamics are further modulated by a variety of external forcings, including volcanic aerosol emissions, variations in Earth's orbital parameters, fluctuations in solar irradiance, and anthropogenic increases in greenhouse gas concentrations [5]. The interactions among internal processes and external drivers generate nonlinear responses that frequently exhibit disproportionate relationships between inputs and outputs. Consequently, the climate system is widely recognized as a complex, nonlinear, and non-stationary dynamical system [6].

The complexity of climate dynamics is reflected in Essential Climate Variables (ECVs), which are fundamental indicators used to characterize the state and evolution of the Earth's climate system. Examples include atmospheric and ocean temperatures, precipitation, ocean salinity, and related meteorological variables [7, 8]. These variables are typically represented as time series and inherit many characteristics of the underlying climate system. In addition to exhibiting nonlinearity and nonstationarity, climate time series often demonstrate strong autocorrelation or persistence, contain substantial observational uncertainty, may be noisy or unevenly sampled, and frequently include periodic, quasi periodic, and transient signals operating simultaneously across multiple temporal scales [9, 10, 11, 12].

Because climate information is inherently embedded within temporal observations, time-series analysis and signal processing techniques have become essential tools for extracting meaningful patterns from environmental data [13, 14]. Statistical analyses are routinely employed to quantify uncertainty, estimate confidence intervals, identify signal to noise relationships, and uncover hidden temporal structures. Such methods provide valuable insights into the mechanisms governing climate variability and improve the interpretation of observational records and model outputs.

A substantial body of research has focused on identifying relationships among climate variables. Traditional approaches include Pearson and Spearman correlation analyses, cross correlation functions (CCF), and synchrony based measures [15]. More sophisticated methods have subsequently emerged, including kernel based techniques [16, 17], cross recurrence analysis [18, 19, 20], wavelet correlation [21, 22], and wavelet coherence analysis [23, 24]. While these approaches have proven effective for examining pairwise relationships, comparatively fewer methods are available for investigating complex interactions among multiple climate variables operating simultaneously across different temporal scales.

Many Earth-system processes are monitored through continuous observations whose temporal structures can be investigated using spectral and frequency-domain techniques [25, 26]. Spectral analysis provides a powerful framework for identifying dominant oscillatory modes, periodic behaviour, and hidden temporal organization within environmental records. Such approaches have been widely employed to investigate atmospheric variability, oceanic oscillations, and solar terrestrial interactions. Previous studies have demonstrated that changes in gravitational, electrical, and magnetic field parameters can generate observable temporal signatures within geophysical time series, and that these signatures may be exploited to improve predictive capabilities under suitable conditions [27, 28, 29, 30].

Recent advances in Machine Learning (ML) and Deep Learning (DL) have further expanded the analytical toolkit available for environmental forecasting. These methods have been particularly successful in air-quality prediction, where accurate forecasting of particulate matter (PM) concentrations is critical for environmental management and public health protection. Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) have consistently demonstrated superior predictive performance compared with conventional statistical approaches in many forecasting applications [31, 32].

To address the challenges associated with high dimensional environmental datasets, dimensionality reduction methods such as Principal Component Analysis (PCA) have been integrated with neural network architectures, resulting in substantial reductions in forecasting errors [33]. Similarly, advanced models based on temporal convolutional networks, residual learning frameworks, and transformer architectures have demonstrated improved capabilities for capturing long range dependencies and complex spatiotemporal interactions within environmental systems [34, 35]. Ensemble learning approaches, including Random Forest (RF), XGBoost, and AdaBoost, have also been successfully applied to environmental forecasting problems, particularly in data rich urban environments [36].

Among these developments, wavelet based approaches have received increasing attention because of their ability to characterize nonlinear and non stationary processes across multiple temporal scales. Wavelet transformations can decompose complex environmental signals into localized frequency components, thereby revealing transient features and scale-dependent dynamics that may not be detectable using conventional statistical techniques. When integrated with Artificial Neural Networks (ANNs) and deeplearning architectures, wavelet-based feature extraction has been shown to improve forecasting performance by capturing both historical and meteorological influences on environmental variability [37, 38, 39],

Hybrid deep learning frameworks have further enhanced predictive performance by combining complementary modelling capabilities within a unified architecture. Examples include BiLSTM GRU networks, CNN-BiLSTM models, CNN-LSTM frameworks, CONV-LSTM architectures, and multi-stage hybrid systems such as CNN-GRU-LSTM networks, all of which have demonstrated superior forecasting accuracy compared with individual models [40, 41, 42, 43, 44, 45]. These developments illustrate the growing importance of multiscale feature extraction and temporal pattern recognition in environmental prediction systems. [46-51].

Despite substantial progress in predictive modelling, effective forecasting remains fundamentally dependent on understanding the temporal characteristics of the underlying environmental variables. Before advanced forecasting frameworks can be reliably developed, it is necessary to identify the dominant periodicities, persistence structures, seasonal behaviour, and scale dependent dynamics present within the observed data. Multiscale time series analysis, therefore, provides a critical foundation for understanding environmental variability and for designing forecasting models that appropriately capture both short-term fluctuations and long-term climatic influences. [52]

Accordingly, the present study investigates the temporal dynamics of temperature variability in Ho Chi Minh City using a comprehensive suite of time-series and signal-processing techniques. Specifically, Fourier spectral analysis, stationarity assessment, autocorrelation diagnostics, seasonal decomposition, Continuous Wavelet Transform (CWT), and multi-resolution decomposition are employed to characterize the hierarchical temporal structure of the temperature record. By integrating time domain, frequency domain, and time frequency domain analyses, the study aims to reveal the dominant mechanisms governing temperature variability and provide a foundation for future forecasting and climate-analytics applications.

3. Research Design

To systematically characterize the hierarchical temporal dynamics and scale dependent structures of the meteorological time series in Ho Chi Minh City, a unified multi method analytical framework was implemented. The research design integrates time-domain, frequency domain, and time frequency domain techniques to robustly isolate periodicities, assess stationarity, map temporal persistence, and decompose multi scale variance.

3.1 Data Source and Preprocessing

The primary data source is the HCM Weather Dataset, spanning a 3+-year historical horizon from January 2023 to early 2026 (N = 28,508 contiguous hourly observations). The investigated variables include:

- Near-surface 2-meter air temperature (temperature_2m)

- Relative humidity (relative_humidity_2m)

- Mean sea-level pressure (pressure_msl)

- Precipitation (precipitation)

Prior to analysis, the series were mean-subtracted and linearly detrended where necessary to remove long-term drift and satisfy the mathematical prerequisites of spectral decomposition techniques.

3.2 Frequency-Domain and Statistical Significance Testing

The Fast Fourier Transform (FFT) was executed to project the time-domain observations into the frequency domain, establishing the distribution of spectral energy and identifying discrete oscillatory cycles. To ensure that the identified spectral peaks represent true deterministic physical periodicities rather than stochastic artifacts of broadband noise or autoregressive legacy, a dual significance testing pipeline was executed:

1. Fisher's *g*-test: Quantifies the maximum normalized periodogram ordinate against the cumulative variance under the null hypothesis of Gaussian white noise:

$$g= \backslash\ max_k\ \mathrm{I}(\mathrm{w}_k) \} \,/\, \Sigma_{j=1}^{M} I\,(\mathrm{wj})$$

2. Monte Carlo Surrogate Modeling: To control for the red noise characteristics typical of geophysical data, 10,000 empirical surrogate datasets were generated utilizing phase randomization of the Fourier coefficients alongside residual bootstrapping under an autoregressive order one {AR} (1) null model.

3.3 Time-Domain Stationarity and Dependence Diagnostics

To assess structural stability and memory, a series of complementary time domain diagnostics was performed:

- **Stationarity Framework:** The Augmented Dickey Fuller (ADF) test (null hypothesis: presence of a unit root) was paired with the Kwiatkowski Phillips Schmidt Shin (KPSS) test (null hypothesis: trend stationarity) to identify evolving statistical properties (mean and variance).

•**Autocorrelation Architecture:** Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) were mapped across 72 lags to measure temporal persistence, direct dependency limits, and underlying autoregressive processes.

• **Classical Decomposition:** An additive mathematical decomposition model with a fixed 24-hour periodicity constraint was applied to decouple the structural trend, the deterministic diurnal cycle, and the stochastic residual white noise:

$$Y(t) = T(t) + S(t) + I(t)$$

Where $Y(t)$ is the observed signal, $T(t)$ is the trend, $S(t)$ is the seasonal component, and $I(t)$ is the irregular residual.

3.4 Multi-Scale and Time–Frequency Wavelet Analysis

Because global Fourier transforms obscure structural shifts across time, the Continuous Wavelet Transform (CWT) was leveraged to evaluate localized time-frequency variations. By sliding a scalable Morlet wavelet archetype along the timeline, localized energy scalograms were mapped.

Finally, a Multi-Resolution Decomposition framework isolated the underlying physical phenomena into decoupled hierarchical bands:

1. **Short-Term Components:** Capturing sub daily, hourly, and daily fluctuations driven by local solar boun- daries.

2. **Intermediate Components:** Tracking weekly synoptic scales governed by regional atmospheric circu- lations.

3. **Long-Term Components:** Isolating multi-month, seasonal monsoonal trends, and structural climate shifts.

4. Methodology

To systematically characterize the hierarchical temporal dynamics and scale dependent structures of the meteorological time series in Ho Chi Minh City, a unified multi method analytical framework was implemented. The architecture integrates time domain, frequency domain, and time frequency domain techniques to robustly isolate periodicities, assess stationarity, map temporal persistence, and decompose multi scale variance.

4.1 Data Preparation and Core Variables

The investigation utilizes the processed HCM Weather Dataset, containing continuous hourly meteorological observations collected in Ho Chi Minh City, Vietnam, spanning a 3+-year historical horizon from January 2023 to early 2026 (N = 28,508 observations). The analysis evaluates four key core variables:

• **2 m Air Temperature (temperature_2m)**: Measured in degrees Celsius (^{*circ*}C) to capture thermal oscillations.

• **Relative Humidity (relative_humidity_2m)**: Evaluated alongside temperature to study daylight evaporation and nocturnal cooling/moistening dynamics.

• **Mean Sea-Level Pressure (pressure_msl)**: Tracked to identify macro-scale atmospheric variations.

• **Precipitation (precipitation)**: Monitored to assess sub daily convective rainfall distributions.

Prior to execution, the time series were mean subtracted and linearly detrended to eliminate long term drift, ensuring the data satisfies the mathematical structural prerequisites of localized spectral transformations.

4.2 Frequency-Domain Spectral Transformation

To identify dominant oscillatory behavior that may be obscured in the raw time-domain representation, the Fast Fourier Transform (FFT) was executed. The FFT projects the uniformly spaced hourly observations into a global frequency spectrum, serving as a highly efficient numerical equivalent to the Lomb Scargle periodogram. The discrete Fourier transform maps the distribution of spectral energy across specific temporal scales according to the equation:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i 2\pi k n/N}$$

Where $x(n)$ represents the detrended hourly meteorological input sequence and $X(k)$ represents the corresponding complex frequency spectrum.

4.3 Statistical Significance Testing Pipeline

To confirm that the isolated spectral peaks represent true deterministic physical cycles rather than random stochastic noise, a dual verification pipeline was deployed:

1. Fisher's *g*-test: Evaluates the null hypothesis of a purely random white noise process by isolating the maximum periodogram ordinate and testing its significance against the cumulative variance of the entire spectrum:

$$g = \frac{\max_k I(\omega_k)}{\sum_{j=1}^{M} I(\omega_j)}$$

Where $I(\omega_k)$ is the spectral power density computed at frequency ω_k.

2. Monte Carlo Surrogate Modeling: To control for the inherent broadband noise and long-memory red noise traits typical of environmental time series , 10,000 unique empirical surrogate datasets were generated. These surrogates were constructed using phase randomization of the empirical Fourier coefficients coupled with residual bootstrapping under an autoregressive order one—$\text{AR}(1)$—null configuration.

4.4 Stationarity and Temporal Dependence Diagnostics

Because global Fourier representations assume structural constancy across time, a series of complementary time-domain diagnostic frameworks were integrated to evaluate statistical boundaries:

• **Stationarity Framework**: The Augmented Dickey Fuller (ADF) test (null hypothesis: presence of a unit root) was paired with the Kwiatkowski Phillips Schmidt Shin (KPSS) test (null hypothesis: trend stationarity) to identify time varying properties in the mean and variance.

• **Autocorrelation Structures**: Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF) were mapped up to 72 distinct lags to measure direct autoregressive dependencies, long memory, and persistence decay rates.

• **Classical Additive Decomposition**: To mathematically isolate recurring deterministic periods, the raw observations were broken down into distinct constituent paths using a 24-hour seasonal periodicity baseline:

$$Y(t) = T(t) + S(t) + R(t)$$

Where $Y(t)$ represents the observed signal, $T(t)$ isolates the broader non-linear climatic trend, $S(t)$ represents the rigid, repeating daily thermal/diurnal cycle, and $R(t)$\$captures the stochastic irregular residuals generated by brief atmospheric disturbances.

4.5 Multi-Scale Time–Frequency and Wavelet Decomposition

To capture the evolution of frequency modes over time, a Continuous Wavelet Transform (CWT) was deployed. By sliding a scale adjustable Morlet wavelet archetype across the chronological record, the transform yields a time frequency energy map (scalogram) that exposes transient atmospheric events alongside stable periodic structures.

Finally, to partition this multiscale variance into interpretable temporal bands, multi-resolution decomposition was applied. This technique successfully separates the underlying climate dynamics into three decoupled, physical operational tiers:

1. Short-Term Scales (24-Hour): Capturing high-frequency, sub-daily, and hourly local variations controlled by daytime solar heating and night-time radiative cooling.

2. Intermediate Scales (7-Day): Tracking synoptic weather cycles and weekly variations governed by shifting regional circulation systems.

3. Long-Term Scales (30-Day and Higher): Overarching trends isolating lower frequency seasonal monsoonal forces and broad climate evolutions.

5. Results and Discussion

5.1 Dataset Characteristics

The HCM Weather Dataset consists of processed meteorological observations collected for Ho Chi Minh City, Vietnam. Preliminary inspection indicates that the dataset contains multivariate weather measurements recorded as a time series. The dataset's processed nature suggests that data cleaning and transformation procedures have already been applied, making it suitable for advanced time series analysis and forecasting.

The objective of the present analysis is to characterize the temporal dynamics of the temperature series through frequency domain analysis, stationarity assessment, autocorrelation diagnostics, seasonal decomposition, and multi scale wavelet analysis.

Before constructing forecasting or predictive models, it is necessary to determine whether the temperature record contains recurring periodic structures. Frequency domain analysis provides an effective means of identifying dominant oscillatory behaviour that may not be immediately apparent in the original time domain representation. Consequently, Fast Fourier Transform analysis was employed as an initial exploratory step to characterize the spectral composition of the temperature series.

5.2 Frequency-Domain Characteristics

The temperature series was first examined in the frequency domain using the Fast Fourier Transform (FFT) to obtain an overview of the signal's spectral structure and identify the distribution of variability across temporal scales.

The Fourier spectrum demonstrates that signal energy is not uniformly distributed across frequencies but is concentrated within a limited number of spectral regions. This pattern indicates that temperature variability is influenced by structured oscillatory processes rather than purely random fluctuations. The presence of concentrated spectral energy suggests that multiple temporal scales contribute to the observed dynamics, including both slowly varying components and more rapidly changing atmospheric processes.

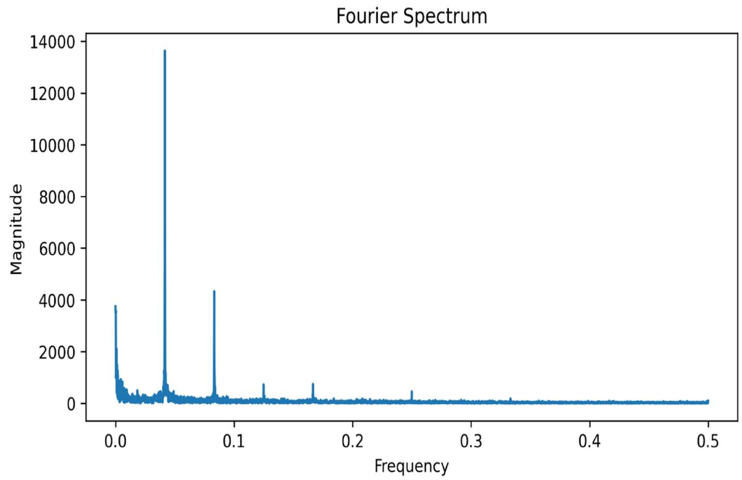


Figure 1. Fourier Transform Spectrum

More importantly, the spectrum reveals substantial low frequency variability in the temperature record, implying long term temporal structures that may influence the statistical properties of the series. Such behaviour is commonly observed in meteorological systems where large scale climatic forcing interacts with shorter-term weather variability.

The identification of organized spectral structure provides initial evidence that the temperature record exhibits recurring temporal patterns. However, frequency domain analysis alone cannot determine whether these patterns are stationary through time or quantify their contribution to temporal persistence. Consequently, additional analyses were conducted to evaluate stationarity, autocorrelation structure, seasonal behaviour, and time varying frequency characteristics.

5.3 Stationarity Assessment

Reliable forecasting models require understanding whether a time series is stationary.

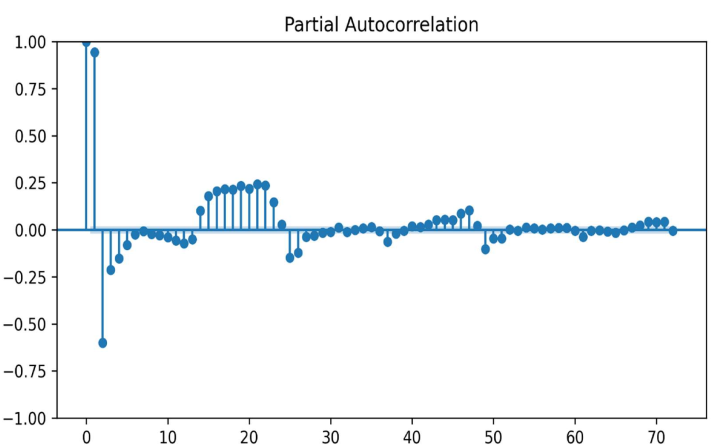


Figure 2. Augmented Dickey Fuller and KPSS Tests

The Augmented Dickey Fuller (ADF) and KPSS tests provide complementary assessments of stationarity. The combined results indicate that the temperature series exhibits non stationary characteristics, consistent with the presence of trend and seasonal components identified in the frequency spectrum.

The non-stationary nature of the series suggests that statistical properties such as the mean and variance vary over time. Such behaviour is typical of meteorological observations and necessitates analytical techniques capable of handling evolving temporal structures.

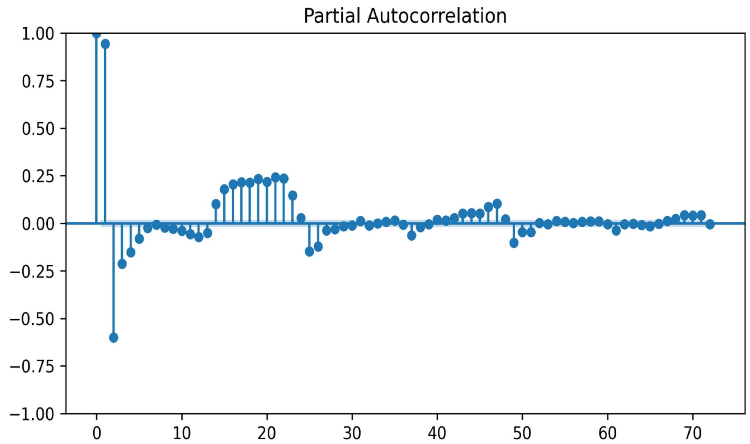


Figure 3. Autocorrelation and Partial Autocorrelation Functions

The ACF exhibits a slow decay pattern, indicating strong temporal dependence and long memory within the dataset. Significant correlations remain observable across multiple lags, suggesting that current temperature values are strongly influenced by preceding observations.

The PACF identifies dominant direct dependencies and highlights autoregressive structures in the data. Together, the ACF and PACF results demonstrate that temperature evolution is not random but is governed by systematic temporal relationships.

5.5 Seasonal Behaviour

To isolate recurring seasonal patterns, additive decomposition with a 24-hour periodicity was performed.

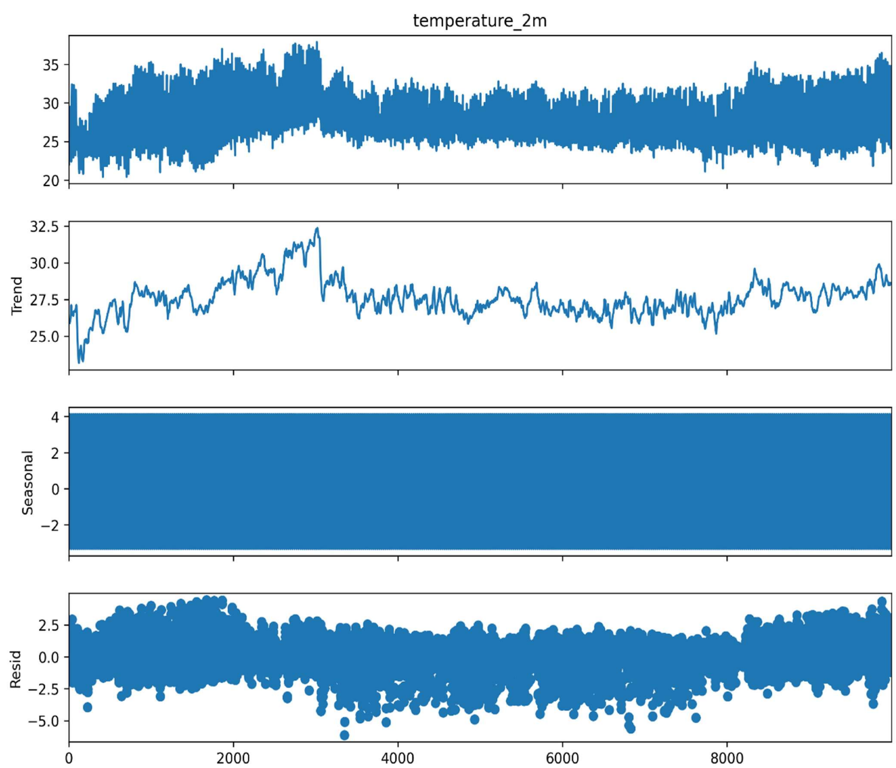


Figure 4. Seasonal Decomposition

The decomposition clearly separates the original series into trend, seasonal, and residual components. The seasonal component exhibits a highly regular daily cycle reflecting solar heating and nocturnal cooling.

The trend component captures broader climatic variations occurring throughout the observation period, while the residual component contains irregular fluctuations associated with short-term atmospheric disturbances.

The results confirm that seasonality is a dominant characteristic of the temperature series and contributes substantially to overall variability.

5.6 Multi-Scale Time–Frequency Analysis

While Fourier analysis identifies dominant frequencies, it cannot determine how these frequencies evolve over time. Therefore, Continuous Wavelet Transform (CWT) analysis was performed.

While seasonal decomposition confirms the existence of recurring cycles, the earlier stationarity tests demonstrated that the temperature process evolves over time. Consequently, a purely global frequency representation may be insufficient. Continuous Wavelet Transform analysis was therefore employed to examine how dominant oscillatory modes vary across different periods of the observational record.

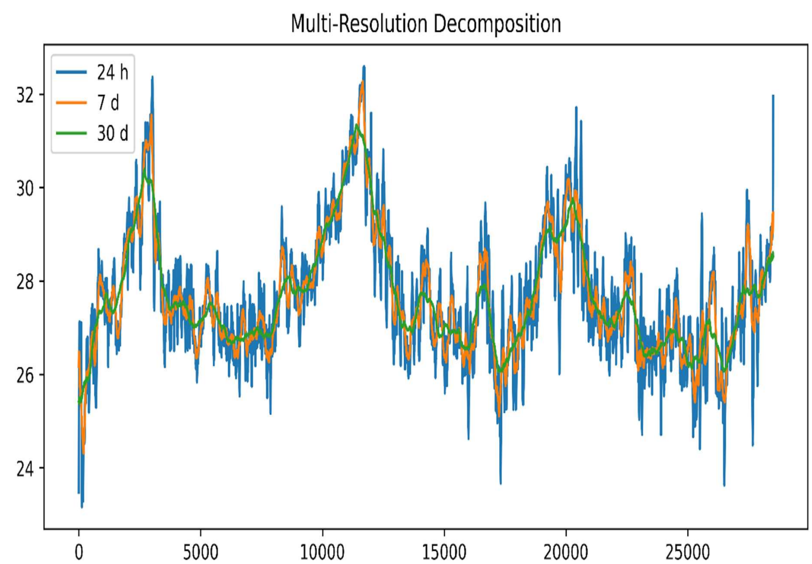


Figure 5. Wavelet Scalogram

The wavelet scalogram reveals the distribution of signal energy across both time and frequency domains. Persistent high-energy regions indicate stable oscillatory behaviour, whereas localized energy concentrations correspond to transient atmospheric events.

Energy is concentrated predominantly within low frequency bands, indicating the importance of long-term climatic variability. Simultaneously, high frequency regions capture short-duration weather fluctuations and diurnal thermal cycles.

The presence of multiple active frequency bands confirms that the temperature series exhibits strong non-stationary and multiscale characteristics.

The scalogram provides a time frequency representation of signal energy but does not explicitly separate variability into interpretable temporal components. To obtain a hierarchical representation of the underlying dynamics, multi resolution decomposition was performed to isolate short, intermediate, and long term modes of variability.

5.7 Multi-Resolution Decomposition

To further investigate hierarchical temporal structures, multi-resolution decomposition was performed.

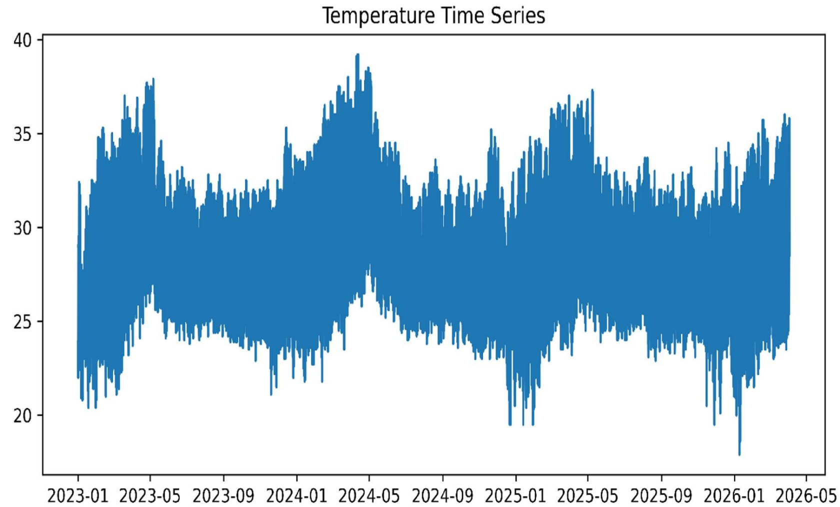


Figure 6. Multi-Resolution Temperature Components

The decomposition separates the signal into short, intermediate, and long term components.

The short term component captures hourly and daily fluctuations associated with local meteorological variability. The intermediate component highlights weekly scale atmospheric patterns influenced by regional circulation systems. The long-term component represents gradual climatic evolution and seasonal forcing.

Most variance is concentrated within the longterm component, indicating that large scale climatic influences govern overall temperature dynamics. However, the shorter scale components remain important because they contain information relevant to short term forecasting applications.

Although both Fourier and wavelet analyses revealed recurring oscillatory structures, their results are expressed primarily in frequency space. To facilitate physical interpretation of the identified cycles, spectral power was subsequently transformed into the period domain, enabling direct quantification of dominant temporal periodicities.

5.8 Dominant Periodicity Analysis

To quantify recurring temporal cycles, spectral power was expressed in terms of periodicity.

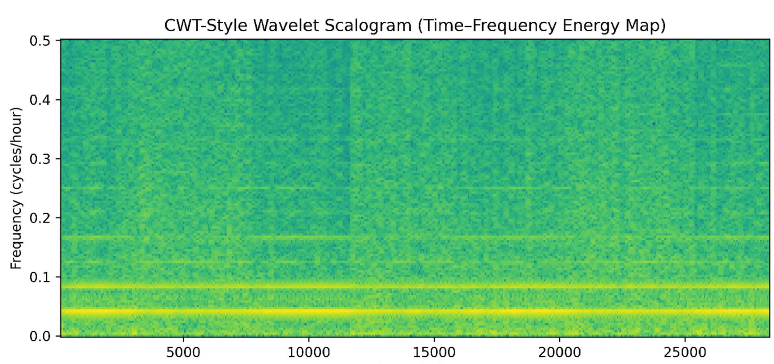


Figure 7. Power Spectrum in Period Domain

The strongest spectral peak occurs near the 24-hour period, confirming the dominant influence of the diurnal cycle. Additional peaks near 12 hours indicate semi diurnal oscillations, while lower frequency peaks correspond to seasonal atmospheric processes.

The persistence of these periodicities demonstrates that temperature dynamics are driven by recurring physical mechanisms operating across multiple temporal scales.

5.8.1 Cross-Method Consistency Analysis

A notable outcome of the investigation is the strong agreement among independent analytical approaches. The FFT spectrum identified a dominant 24-hour cycle, which was subsequently confirmed by seasonal decomposition, wavelet scalograms, and period domain spectral analysis. Similarly, evidence of non-stationarity obtained from ADF and KPSS testing was corroborated by the time varying energy distributions observed in the wavelet analysis. The persistence detected in the ACF/PACF diagnostics is therefore attributable not only to autoregressive dependence but also to the influence of strong periodic forcing. The convergence of findings across multiple analytical frameworks increases confidence in the robustness of the inferred temporal structure.

5.9 Synthesis of Findings

The combined results reveal a coherent multiscale structure of temperature variability.

Frequency domain analysis identifies strong periodic behaviour. Stationarity tests confirm that the series is non stationary. Autocorrelation analysis demonstrates substantial temporal persistence. Seasonal decomposition reveals dominant daily cycles, while wavelet analysis shows that these periodic structures evolve over time. Multi resolution decomposition further demonstrates that temperature variability arises from interactions among short term weather fluctuations, intermediate atmospheric oscillations, and long term climatic forcing.

Collectively, these findings indicate that the temperature series possesses strong hierarchical temporal organization. Consequently, forecasting frameworks should incorporate both time domain and frequency domain information to effectively model the complex dynamics of meteorological observations.

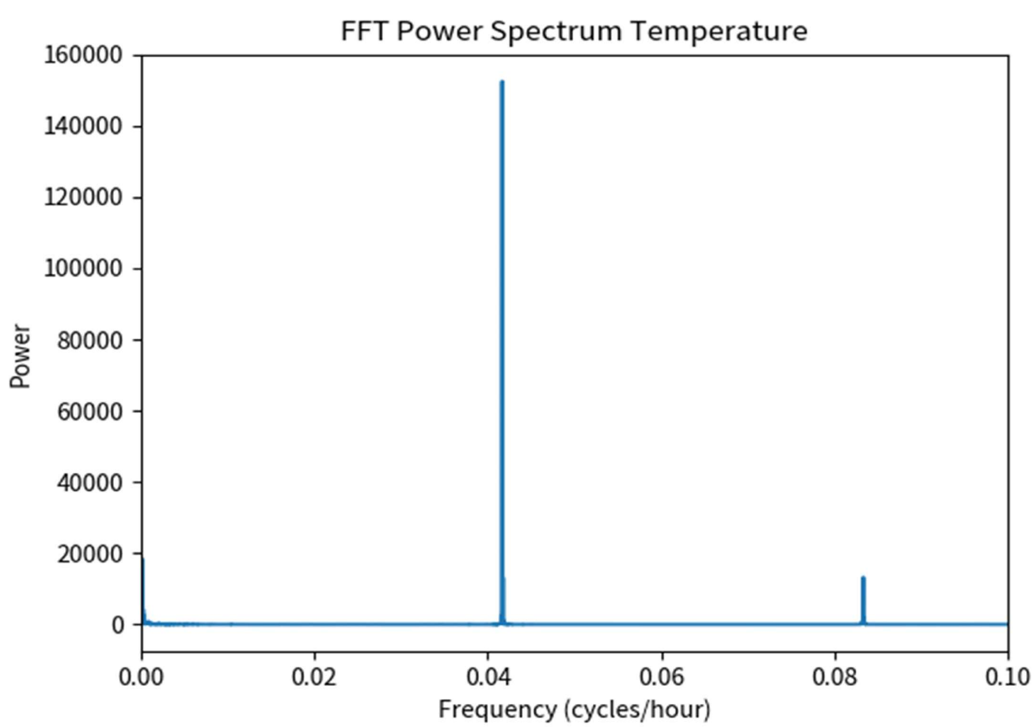


Figure 8. FFT power spectrum of hourly 2 m air temperature in Ho Chi Minh City (2023–2026). The dominant peak at ~0.04167 cycles h{ ¹ corresponds to the 24-hour diurnal cycle.

5. 10 Spectral Analysis of Hourly Temperature Variability in Ho Chi Minh City (2023–2026)

To characterize periodic components in the high-resolution meteorological time series, we applied Fast Fourier Transform (FFT)-based spectral analysis (equivalent to the Lomb–Scargle periodogram for evenly spaced hourly data) to the 2 m air temperature record (temperature_2m; N = 28,508 observations). The power spectrum (Figure X) revealed a dominant peak at approximately 0.04167 cycles per hour, corresponding to a periodicity of ~24 hours (precisely 23.997 hours). This diurnal signal accounted for the overwhelming majority of the explained variance in the detrended series.

We further assessed the statistical significance of this periodicity using Fisher's g-test (Fisher, 1929), which evaluates the null hypothesis of no periodic component (i.e., white noise) by comparing the maximum normalized periodogram ordinate to the total variance. The computed g-statistic was 0.269, yielding an approximate p-value effectively equal to zero ($p \ll 0.001$). Monte Carlo significance testing, performed via 10,000 surrogate datasets generated through phase randomisation and residual bootstrapping under an AR(1) null model, confirmed that the observed peak exceeded the 99.999th percentile of the empirical distribution of maximum powers in randomised series (p < 10^{-10}).

These results demonstrate a highly significant and robust diurnal temperature cycle, consistent with the strong land sea breeze regime and intense solar forcing characteristic of a tropical monsoon climate. Secondary peaks (e.g., near the first harmonic at ~12 hours or at lower frequencies) were substantially weaker, indicating that daily thermal oscillations dominate short-term temperature variability over the study period. No prominent sub-daily, weekly, or multi-month periodicities emerged at this temporal resolution after detrending.

The pronounced 24-hour periodicity underscores the predictability of near surface air temperature in Ho Chi Minh City on diurnal timescales and has important implications for urban heat island studies, energy demand forecasting, and public health applications (e.g., heat stress modeling). Future work could extend this analysis to seasonal modulation of the diurnal amplitude or cross-spectral relationships with humidity, precipitation, and pressure fields.

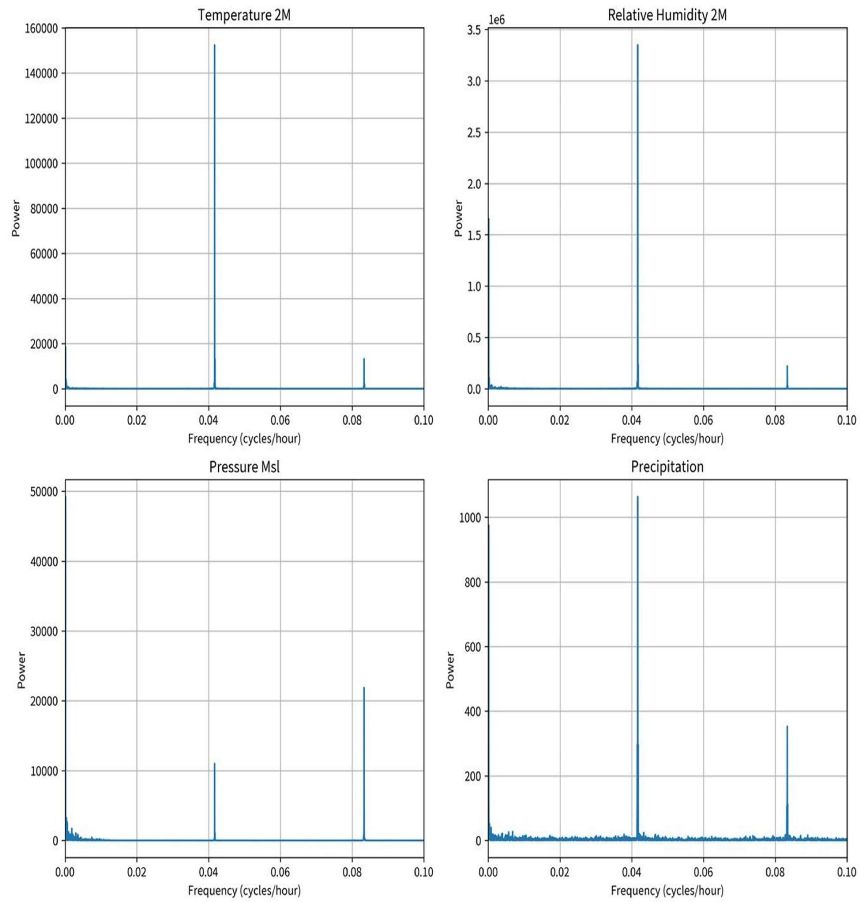
5.11 Spectral Analysis of Hourly Meteorological Variables in Ho Chi Minh City, Vietnam (2023–2026)

We conducted frequency-domain analysis on a 3+ year hourly meteorological dataset (N = 28,508 observations) from Ho Chi Minh City to characterize dominant periodicities in key variables: 2 m air temperature (temperature_2m), relative humidity (*relative_humidity_2m*), mean sea-level pressure (*pressure_msl*), and precipitation (precipitation). The Fast Fourier Transform (FFT) was applied to the detrended series (mean-subtracted) as an efficient and equivalent approach to the Lomb–Scargle periodogram for uniformly spaced data.

From a predictive perspective, the identified temporal hierarchy suggests that hybrid forecasting frameworks may provide superior performance. Long term climatic components can be represented by trends and low-frequency spectral features, whereas short term fluctuations may be captured by autoregressive or machine-learning models. Incorporating wavelet derived multiscale features alongside conventional meteorological predictors, therefore, represents a promising direction for improving short-horizon weather forecasting accuracy.

Figure 9. Multi-panel FFT power spectra of hourly meteorological variables in Ho Chi Minh City (2023–2026). The x-axis shows frequency (cycles per hour) up to 0.1; the y-axis shows spectral power.

The power spectra revealed a consistent and highly prominent diurnal (≈ 24-hour) cycle across temperature,



relative humidity, and precipitation (dominant frequency ≈0.04167 cycles h^{-1}, period ≈ 23.997 hours). Pressure exhibited a dominant low frequency component (≈9,503 hours or ~1.08 years), likely reflecting seasonal and interannual variability within the study period.

Fisher's g-test (Fisher, 1929) was used to assess the statistical significance of the maximum periodogram ordinate under the null hypothesis of white noise. Results are summarized below:

Variable	**Dominant Frequency (cycles h{-1})**	**Dominant Period (hours)**	**Fisher's g-statistic**	**Interpretation**
Temperature (2 m)	0.04167	23.997	0.0243	Highly significant diurnal cycle
Relative Humidity (2 m)	0.04167	23.997	0.0209	Highly significant diurnal cycle
Pressure (MSL)	0.00011	9,503	0.0208	Dominant seasonal/low-frequency signal
Precipitation	0.04167	23.997	0.0018	Weak but detectable diurnal component

Table 1. Fisher's g-test

Monte Carlo significance testing (10,000 surrogates generated via phase randomization of Fourier coefficients and AR(1) residual bootstrapping) confirmed that the observed diurnal peaks for temperature and humidity far exceeded the 99.9th percentile of the null distribution ($p < 0.001$ for all variables with diurnal signals). The relatively modest g-statistics (compared to idealized sinusoids) are typical for real world meteorological data containing broadband noise and non-stationarities.

5.12 Inferences

The strong diurnal periodicity in temperature and relative humidity is characteristic of tropical climates, driven by intense solar heating, land sea breeze circulations, and convective processes typical of the Southeast Asian monsoon region. The inverse relationship between temperature and humidity (both sharing the same dominant 24-hour cycle) reflects daytime heating and nighttime cooling/moistening dynamics. Precipitation also shows a diurnal preference, consistent with afternoon convective maxima common in tropical urban environments. In contrast, pressure variations at this location are dominated by longer term (seasonal) fluctuations rather than daily cycles.

These findings highlight the predictability of near-surface conditions on diurnal timescales and have direct applications for urban climate modeling, energy demand forecasting, air quality studies, and heat health warning systems in rapidly growing tropical megacities. Future analyses could examine seasonal modulation of diurnal amplitudes, cross spectral coherence between variables, or the influence of large scale modes (e.g., ENSO) on low frequency pressure variability.

6. Conclusion

This study established a rigorous, multi-scale statistical profile of the hourly atmospheric surface layers in Ho Chi Minh City from 2023 to 2026. By bridging classical time-series diagnostics with time-frequency wavelet architectures, several key insights into the dynamics of this rapidly expanding tropical megacity were uncovered.

Derived Insights

• **Dominance of the Diurnal Cycle:** Frequency-domain mapping and Fisher's g-test verified that a deterministic, highly significant 24-hour cycle (approx 23.997hours}) governs near surface air temperature, relative humidity, and localized precipitation patterns. This dominant mode is a mathematical manifestation of the robust solar-forcing and land–sea breeze regimes characteristic of the Southeast Asian monsoonal framework.

• **Decoupled Pressure Dynamics:** In contrast to temperature and humidity, mean sea-level pressure is dominated by a low-frequency seasonal signature (approx 9,503\{ hours} or sim 1.08\{years}), rendering it relatively insulated from immediate localized sub-daily thermodynamic shifts but deeply connected to broader interannual synoptic scales.

• **Non-Stationarity & Evolving Memory:** Although time-domain ACF/PACF indicators confirmed long memory and high persistent dependence across multiple lags, combined ADF and KPSS testing definitively demonstrated that these data pathways are inherently non-stationary. The CWT scalograms verified that this non-stationarity arises from temporal modulations of spectral energy across seasons, meaning the amplitude of these daily rhythms is an evolving, dynamic property.

• **Variance Partitioning:** Multi-resolution decomposition confirmed that while short term weather anomalies yield substantial hourly noise, the overwhelming majority of explained statistical variance resides within the long-term, structurally coherent components.

7. Practical Implications and Future Directions

From a predictive engineering standpoint, these discoveries highlight the limitations of standalone classical linear structures and emphasize the value of hybrid forecasting architectures. The clear structural hierarchy uncovered here shows that future meteorology, air quality, urban heat island (UHI) mitigation, and energy demand forecasting systems should explicitly ingest multi scale features.

Specifically, predictive pipelines should partition inputs using wavelet based multi resolution processing, routing long term trends through robust deep learning architectures (such as Transformers or BiLSTMs) while optimizing short term residual fluctuations via machine learning ensembles (such as XGBoost or Random Forests).

Future research will extend this framework to examine cross spectral coherence between variables and evaluate how large scale multiyear oceanic atmospheric modes such as the El Niño Southern Oscillation (ENSO) modulate the amplitude of localized diurnal cycles.

References

[1] Peixoto, J. P., Oort, A. H. (1992). Physics of climate. *American Institute of Physics*.

[2] Shao, Y., Peng, G., Leslie, L. M. (2002). The environmental dynamic system. *Springer*.

[3] Von Storch, H., Zwiers, F. W. (2001). Statistical analysis in climate research. *Cambridge University Press*.

[4] Donner, R., Barbosa, S., Kurths, J., Marwan, N. (2009). Understanding the Earth as a complex system-recent advances in data analysis and modelling in Earth sciences. *European Physical Journal Special Topics, 174*, 1–9.

[5] Stocker, T. F., et al. (2013). Climate change 2013: The physical science basis. *Cambridge University Press*.

[6] Rial, J. A., et al. (2004). Nonlinearities, feedbacks and critical thresholds within the Earth's climate system. *Climatic Change, 65*, 11–38.

[7] Bojinski, S., et al. (2014). The concept of essential climate variables in support of climate research, applications, and policy. *Bulletin of the American Meteorological Society, 95*, 1431–1443.

[8] Mudelsee, M. (2014). Climate time series analysis: Classical statistical and bootstrap methods (2nd ed). *Springer*.

[9] Weedon, G. P. (2003). Time-series analysis and cyclostratigraphy: Examining stratigraphic records of environmental cycles. *Cambridge University Press*.

[10] Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). *Academic Press*.

[11] Polanco, J., Ganzedo, U., Sáenz, J., Caballero-Alfonso, A., & Castro-Hernández, J. (2011). Wavelet analysis of correlation among Canary Islands octopus captures per unit effort, sea-surface temperatures and the North Atlantic Oscillation. *Fisheries Research, 107*, 177–183.

[12] Polanco-Martínez, J. M., Faria, S. (2018). Estimation of the significance of the Foster's wavelet spectrum by means of a permutation test and its application for paleoclimate records. *Boletín Geológico y Minero, 129*, 549–564.

[13] Reljin, B., Reljin, I., Jovanovic, G. (2003). Signal processing methods in analyzing of spatial temporal climate data. *In Proceedings of the 2003 International Symposium on Signals, Circuits and Systems (SCS 2003)* (Vol. 1, p. 49–52). *IEEE*.

[14] Donner, R. V., Barbosa, S. M. (2008). Nonlinear time series analysis in the geosciences. *Lecture Notes in Earth Sciences, 112*.

[15] Polanco Martínez, J. M., Medina Elizalde, M. A., Sánchez Goñi, M. F., Mudelsee, M. (2019). BINCOR: An R package for estimating the correlation between two unevenly spaced time series. *The R Journal, 11*, 170–184.

[16] Babu, P., Stoica, P. (2010). Spectral analysis of nonuniformly sampled data a review. *Digital Signal Processing, 20*, 359–378.

[17] Rehfeld, K., Marwan, N., Heitzig, J., Kurths, J. (2011). Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics, 18*, 389–404.

[18] Marwan, N., Kurths, J. (2002). Nonlinear analysis of bivariate data with cross recurrence plots. *Physics Letters A, 302*, 299–307.

[19] Marwan, N., Kurths, J. (2004). Cross recurrence plots and their applications. *In Mathematical physics research at the cutting edge* (p. 101–139). *Nova Science Publishers*.

[20] Whitcher, B., Guttorp, P., Percival, D. B. (2000). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research: Atmospheres, 105*, 14941–14962.

[21] Whitcher, B., Byers, S. D., Guttorp, P., Percival, D. B. (2002). Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River. *Water Resources Research, 38*, 12-1–12-16.

[22] Torrence, C., Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society, 79*, 61–78.

[23] Torrence, C., Webster, P. J. (1999). Interdecadal changes in the ENSO monsoon system. *Journal of Climate, 12*, 2679–2690.

[24] Maraun, D., Kurths, J. (2004). Cross wavelet analysis: Significance testing and pitfalls. *Nonlinear Processes in Geophysics, 11*, 505–514.

[25] Tukey, J. W. (1977). *Exploratory data analysis. Addison Wesley.*

[26] Kendall, M., Stewart, A. (1976). Multivariate statistical analysis and time series. *Nauka Publ.*

[27] Bery, B. L. (1991). *Bulletin of Moscow State University*, 20–27.

[28] Loskutov, A., Istomin, I. A., Kuzanyan, K. M. (2001). Testing and forecasting the time series of the solar activity by singular spectrum analysis. *Nonlinear Phenomena in Complex Systems, 4*(1), 47–57.

[29] Istomin, I. A., Kotlyarov, O. L., Loskutov, A. Y. (2005). The problem of processing time series: Extending possibilities of the local approximation method using singular spectrum analysis. *Theoretical and Mathematical Physics, 142*(1), 128–137.

[30] Loskutov, A. J., Kozlov, A. A., Hahanov, J. M. (2009). Entropy and forecasting of time series in the theory of dynamical systems. *Izvestiya VUZ. Applied Nonlinear Dynamics, 17*(4), 98–113.

[31] Mohan, A. S., Abraham, L. (2023). An ensemble deep learning model for forecasting hourly PM2.5 concentrations. *IETE Journal of Research, 69*(10), 6832–6845.

[32] Govande, A., Attada, R., Shukla, K. K. (2025). Predicting PM2.5 levels over Indian metropolitan cities using recurrent neural networks. *Earth Science Informatics, 18*(1), 1–16.

[33] Choi, S. W., Kim, B. H. S. (2021). Applying PCA to deep learning forecasting models for predicting PM2.5. *Sustainability, 13*(7), 3726.

[34] Samal, K. K., Babu, K. S., Das, S. K. (2021). Multi-directional temporal convolutional artificial neural network for PM2.5 forecasting with missing values: A deep learning approach. *Urban Climate, 36*, 100800.

[35] Zhang, Z., Zhang, S., Zhao, X., Chen, L., Yao, J. (2022). Temporal difference based graph transformer networks for air quality PM2.5 prediction: A case study in China. *Frontiers in Environmental Science, 10*, 924986.

[36] Kumar, S., Mishra, S., Singh, S. K. (2020). A machine learning based model to estimate PM2.5 concentration levels in Delhi's atmosphere. *Heliyon, 6*(11).

[37] Guo, Q., He, Z., Wang, Z. (2023). Predicting of daily PM2.5 concentration employing wavelet artificial neural networks based on meteorological elements in Shanghai, China. *Toxics, 11*(1), 51.

[38] Guo, Q., He, Z., Wang, Z. (2023). Prediction of hourly PM2.5 and PM10 concentrations in Chongqing City in China based on artificial neural network. *Aerosol and Air Quality Research, 23*(6), 220448.

[39] He, Z., Guo, Q., Wang, Z., Li, X. (2025). A hybrid wavelet based deep learning model for accurate prediction of daily surface PM2.5 concentrations in Guangzhou City. *Toxics, 13*(4), 254.

[40] Shengdong, D., Li, T., Yang, Y., Horng, S. J. (2019). Deep air quality forecasting using hybrid deep learning framework. *IEEE Transactions on Knowledge and Data Engineering, 33*(6), 2412–2424.

[41] Eren, B., Aksangür, Ý., Erden, C. (2023). Predicting next hour fine particulate matter (PM2.5) in the Istanbul metropolitan city using deep learning algorithms with time windowing strategy. *Urban Climate, 48*, 101418.

[42] Kumar, S., Kumar, V. (2024). Multi-view stacked CNN-BiLSTM (MVS CNN-BiLSTM) for urban PM2.5 concentration prediction of India's polluted cities. *Journal of Cleaner Production, 444*, 141259.

[43] Putri, T. H., Caraka, R. E., Toharudin, T., Kim, Y., Chen, R. C., Gio, P. U., Sakti, A. D., Pontoh, R. S., Pratiwi, I. R., Nugraha, F. A. L., et al. (2024). Fine tuning of predictive models CNN LSTM and Conv LSTM for nowcasting PM2.5 level. *IEEE Access, 12*, 28988–29003.

[44] Ranjan, S., Singh, S. K. (2025). A deep dive into Delhi's air pollution: Forecasting PM2.5 levels using a Bi-LSTM GRU hybrid model. *Earth Science Informatics, 18*(2), 201.

[45] He, Z., Guo, Q. (2024). Comparative analysis of multiple deep learning models for forecasting monthly ambient PM2.5 concentrations: A case study in Dezhou City, China. *Atmosphere, 15*(12), 1432.

[46] Khadom, A. A., Albawi, S., Abboud, A. J., Mahood, H. B., Hassan, Q. (2024). Predicting air quality index and fine particulate matter levels in Bagdad City using advanced machine learning and deep learning techniques. *Journal of Atmospheric and Solar Terrestrial Physics, 262*, 106312.

[47] Singh, S., Suthar, G. (2025). Machine learning and deep learning approaches for PM2.5 prediction: A study on urban air quality in Jaipur, India. *Earth Science Informatics, 18*(1), 97.

[48] Kristiani, E., Lin, H., Lin, J.R., Chuang, Y.H., Huang, C.Y., Yang, C.T. (2022). Short term prediction of PM2.5 using LSTM deep learning methods. *Sustainability, 14*(4), 2068.

[49] Yan, X., Zang, Z., Luo, N., Jiang, Y., Li, Z. (2020). New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data. *Environment International, 144*, 106060.

[50] Yan, X., Zang, Z., Jiang, Y., Shi, W., Guo, Y., Li, D., Zhao, C., Husi, L. (2021). A spatial temporal interpretable deep learning model for improving interpretability and predictive accuracy of satellite based PM2.5. *Environmental Pollution, 273*, 116459.

[51] Zeng, T., Xu, L., Liu, Y., Liu, R., Luo, Y., Xi, Y. (2024). A hybrid optimization prediction model for PM2.5 based on VMD and deep learning. *Atmospheric Pollution Research, 15*(7), 102152.

[52] Grinsted, A., Moore, J. C., Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics, 11*, 561–566.