A Comprehensive Time Series Analysis of Seasonal and Trend Patterns In Temperature Data of Visakhapatnam

Deep Karan Singh[1] & Nisha Rawat*
[1]India Meteorological Department, MoES, Visakhapatnam, India
*Meteorological Office, INS Dega, Visakhapatnam, India
Email: deep.karan.singh@imd.gov.in

ABSTRACT

This research presents an in-depth analysis of urban temperature data through advanced time series decomposition and statistical modelling, aiming to elucidate the complex dynamics of temperature variations within an urban context. Employing a robust methodology that combines seasonal decomposition, the Augmented Dickey-Fuller test for stationarity, and autocorrelation analyses, the study comprehensively explores both the predictable and random components of temperature fluctuations. Key findings indicate that the trend accurately captures the central tendencies and seasonal patterns of urban temperatures, with most predictions falling within an acceptable range of the actual measurements. The seasonal and trend components of the time series reveal consistent long-term patterns and clear seasonal variations, essential for understanding and forecasting weather changes. Additionally, the analysis of residuals, particularly through Kernel Density Estimation and boxplots, highlights the occurrences of extreme temperature deviations and identifies potential areas for refinement. This study contributes significantly to the fields of urban climatology and meteorological forecasting by providing detailed insights into the microclimatic conditions of urban areas. The findings underscore the importance of understanding temperature variability and extremes, especially in light of changing climate patterns and urban development. Furthermore, the research identifies pathways for future work, emphasizing integrating additional environmental factors, exploring more sophisticated modelling techniques, and enhancing predictive capabilities for extreme weather events. Overall, this research offers valuable implications for urban planners, policymakers, and climate scientists in devising strategies to mitigate the impacts of extreme temperatures and improve urban living conditions.

Keywords: *Urban climatology, time series analysis, temperature variability, statistical modelling, extreme weather events*

1. Introduction

In the era of escalating climate concerns, the analysis of temperature data has never been more critical. Urban areas, in particular, present a unique climatic profile due to the urban heat island (UHI) effect, where modifications in land surfaces and urban structures lead to higher temperatures than rural ones. This study delves into the intricate patterns of urban temperature variations, aiming to decode the underlying seasonal and trend components that characterize these changes over time.

This research is grounded in a comprehensive examination of urban climatology and time series analysis, drawing upon seminal works and contemporary studies. Oke's seminal work (1982) lays the foundation for understanding the UHI effect, providing a detailed examination of its energetic basis. Stewart & Oke (2012) extend this understanding by categorizing urban areas into local climate zones, offering a systematic approach to studying urban temperature variations. These foundational texts are crucial for contextualizing urban heat islands' spatial and temporal dynamics, as observed in the current study.

This research's significance lies in its contribution to the broader field of climatology and its potential applications in urban planning and public health. As cities grow, understanding temperature dynamics becomes essential for sustainable urban development and mitigating heat-related health risks. Studies by Meehl & Tebaldi (2004) and Tan et al. (2010) discuss the increasing intensity and frequency of heatwaves due to climate change, underscoring the health implications in urban settings. These studies provide a backdrop for the importance of accurate temperature modelling and forecasting in mitigating heat-related health risks in urban populations.

Further, research by Grimmond (2007) emphasizes the complexity of urban climates and the need for detailed observational and modelling approaches to understand urban temperature dynamics. Additionally, Li & Bou-Zeid (2013) explored the interaction between urban forms and heat waves, highlighting the vulnerability of specific urban layouts to extreme temperatures. This understanding is critical for informing urban design strategies to reduce heat exposure.

Employing a comprehensive time series analysis, this research leverages advanced statistical methods and visualization tools to dissect the maximum temperature data of Visakhapatnam Airport. By adopting the Seasonal and Trend decomposition of time series (STL) approach, augmented with statistical tests such as the Augmented Dickey-Fuller test, the research provides a nuanced understanding of the temporal patterns in temperature data. This multifaceted analysis is crucial for identifying long-term trends, discerning cyclical patterns, and isolating irregular components within the urban temperature data.

The methodology used in this study is informed by the works of Box et al. (2015), Chatfield (2016), and Hamilton (1994), who offer comprehensive methodologies for time series analysis, forecasting, and control, which are fundamental to the decomposition and analysis of temperature data. Recent advancements in statistical modelling, such as those by Hyndman & Athanasopoulos (2018), further enhance the robustness of the analysis by providing tools for handling complex time series data.

The implications of urban temperature dynamics extend beyond immediate climatic impacts, affecting energy consumption, public health, and urban infrastructure. Akbari, Pomerantz, & Taha (2001) discuss the potential for reflective surfaces and vegetation to mitigate UHI effects, while Sailor (2014) reviews the role of building design in urban temperature management. These studies underscore the importance of integrating climate-responsive strategies into urban planning.

Additionally, the research builds upon the findings of Santamouris (2014) and Stone et al. (2010), who discuss strategies for cooling cities and the relationship between urban form and vulnerability to extreme heat. Their studies highlight the role of urban design and green infrastructure in mitigating UHI effects and adapting to climate change.

The objectives of this study are twofold: firstly, to meticulously analyze the maximum temperature data to extract meaningful insights about its trend, seasonality, and irregular components; and secondly, to interpret these findings in the context of urban climatic changes, offering a valuable perspective for policymakers, urban planners, and environmental researchers. By doing so, this paper aims to contribute a detailed data-driven narrative to the discourse on urban climate resilience and adaptation strategies.

Through this investigation, the research aims to bridge the gap between raw meteorological data and actionable knowledge, contributing to the growing body of literature on urban climate analysis and environmental monitoring. It endeavours to provide a comprehensive analysis that not only elucidates the complex dynamics of urban temperatures but also examines the implications of these patterns for urban dwellers and the environment. Furthermore, this study seeks to inspire future research and action, encouraging a proactive approach to understanding and managing the climatic challenges posed by urban environments.

2. Data and Methodology

This study employs a multifaceted approach to analyze urban temperature data through advanced time series analysis techniques, integrating various statistical models and tests to ensure a comprehensive understanding of the underlying patterns. The methodology is meticulously crafted to dissect and interpret the complex patterns in temperature data, focusing on trend, seasonality, and residual components.

Initially, the data collection and preprocessing phase involves aggregating daily maximum temperature readings of Visakhapatnam airport from 1969 to 2023, ensuring data integrity and relevance. The dataset's 'Date' column is transformed into datetime format and indexed accordingly, vital for chronological data analysis. This transformation is represented by the equation:

$$Ti = f\,(Di) \qquad \text{……(I)}$$

Where:

Ti is the temperature reading and Di is the corresponding date.

For the time series decomposition, the Seasonal and Trend decomposition using Loess (STL) approach is applied. This technique is expressed as:

$$Yt = Tt + St + Rt \qquad \text{……(II)}$$

Where

- Yt represents the observed data,
- Tt represents the trend component,
- St represents the seasonal component,
- Rt represents the residual component at time t.

The decomposition is tailored to an annual cycle, considering the Earth's revolution around the sun, thereby setting the decomposition period to 365 days. The Seasonal and Trend decomposition using Loess (STL) assumes that the time series can be decomposed into additive components—trend, seasonality, and residuals. This method is highly flexible and allows for the variation in seasonal patterns over time. However, it assumes that the seasonality is periodic and consistent, which may not capture all forms of seasonality in the data, particularly if there are irregular or complex seasonal cycles. Additionally, STL requires a predefined seasonal cycle length (in our case, 365 days) that may not fully capture the nuances of temperature variations that occur over shorter or irregular periods. The STL approach is selected for its flexibility and robustness in handling various types of seasonal patterns and its ability to fit the trend component adaptively, providing a more refined analysis of the temperature data. In the statistical analysis phase, the Augmented Dickey-Fuller (ADF) test is conducted to assess stationarity, which is crucial for the reliability of subsequent analyses.

The Augmented Dickey-Fuller (ADF) test assumes that the series under analysis is either stationary or can be made stationary through transformation. However, the ADF test might be sensitive to the presence of structural breaks or long-term cycles in the data, which could lead to incorrect conclusions about stationarity.

The ADF test hypothesis can be formulated as:

H_0: The series has a unit root (non-stationary) ……(III)

H_1: The series does not have a unit root (stationary) ……(IV)

A non-stationary series would indicate that the statistical properties of the data change over time, which could bias the analysis. In the event of non-stationarity, transformation techniques such as differencing or logarithmic transformation may be applied to stabilize the variance and mean.

Autocorrelation and partial autocorrelation functions (ACF and PACF) are then utilized to identify any autocorrelation in the residuals. These are represented as:

$$ACF\ (k) = \frac{\sum_{t=1}^{n-k}(Y_t-\bar{Y})(Y_{t+k}-\bar{Y})}{\sum_{t=1}^{n}(Y_t-\bar{Y})^2} \quad \text{….(V)}$$

$$PACF\ (k) = Corr\ (Y_t, Y_{t+k} \mid Y_{t+1}, \dots, Y_{\ t+k-1}) \quad \text{……(VI)}$$

where:

- k is the lag
- Yt is the value of the time series at time t
- ‾Y is the mean of the series.

The analysis of autocorrelation (ACF) and partial autocorrelation (PACF) functions is fundamental in identifying the presence of autoregressive or moving average processes in the residuals. However, these functions can sometimes be difficult to interpret, particularly in complex data structures. These analyses have been applied conservatively, ensuring that identified patterns are robust and not artefacts of noise or model misspecification.

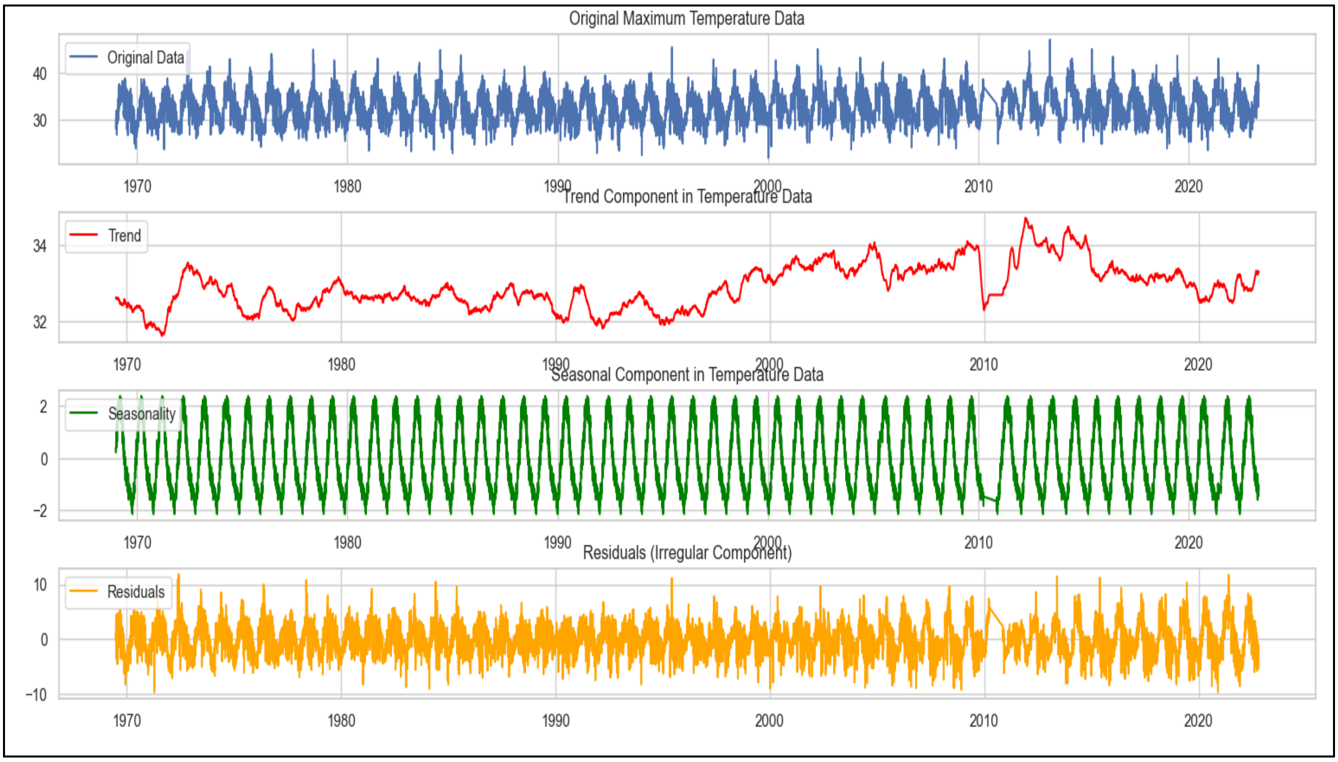

Figure 1: Analysis of temperature data.

Visualization plays a pivotal role in this study, employing statistical graphing tools to present the data, trends, and patterns effectively.

The meteorological analysis dives into each decomposed component, exploring the mean and standard deviation of the trend component, analyzing seasonal patterns, and investigating residuals for anomalies. This in-depth analysis is supported by additional statistical tests, including tests for homoscedasticity and normality in the residuals, providing a deeper understanding of the data's properties and the validity of the decomposition. The additional statistical tests for homoscedasticity and normality in residuals are crucial for validating the underlying assumptions. However, these tests themselves have limitations, particularly when dealing with large datasets where even minor deviations from normality or homoscedasticity may be flagged as significant. These have been addressed by complementing statistical tests with visual diagnostics to ensure that the results are not only statistically significant but also meaningful in a practical context.
This approach interweaves statistical rigour with practical data handling and sophisticated visualization techniques, forming a robust, comprehensive urban temperature data analysis framework. This approach is designed to be adaptable, allowing for its application to diverse urban contexts and datasets, thereby extending its utility beyond the scope of this singular study. While the methodology employed is designed to be adaptable to different urban contexts, we acknowledge that the results and insights derived from this study are specific to the dataset used and may not be directly applicable to other regions with different climatic conditions. Furthermore, the methodology incorporates a feedback loop, where preliminary findings are reviewed, and the analysis is iteratively refined to ensure the most accurate and insightful results. Through this rigorous and detailed approach, the study aims to provide a deep and nuanced understanding of urban temperature dynamics, contributing valuable insights to the field of urban climatology.

3. Results

The results of the time series analysis of the maximum temperature data of Visakhapatnam airport reveal distinct patterns in the original data, trend, seasonality, and residuals. These findings are visually represented through a series of plots (Fig 1), each highlighting a different component of the temperature data.

The plot of the original data (Fig 1a) provides a comprehensive view of the maximum temperatures over the observed period (1969-2023). It showcases the raw temperature readings without any modifications or smoothing, allowing for the visualization of both short-term fluctuations and long-term trends. The variations in the original data reflect the complex interplay between natural climatic variability, such as monsoon cycles, and anthropogenic influences, such as urbanization, which contribute to the urban heat island effect. This baseline plot is crucial for understanding the overall context of the decomposed components.

The trend component (Fig 1b) illustrates the long-term progression or direction in the temperature data, effectively smoothing out short-term fluctuations to reveal the underlying movement over time. This component is crucial for understanding how the maximum temperatures have evolved. The trend line indicates an overall temperature increase, reflecting broader climatic changes such as global warming. The observed trend is particularly significant for urban planners and public health officials, as it suggests a sustained rise in temperatures that could exacerbate heat-related health risks and increase energy demands for cooling.

The seasonal component (Fig 1c) captures the recurring patterns or cycles in the temperature data that occur within a fixed period - in this case, annually. This component is critical for understanding how temperatures fluctuate seasonally, closely tied to the monsoon cycle and other regional climatic patterns. The seasonal plot is essential for understanding the rhythmic nature of temperature changes and can be used to predict future patterns based on historical data.

The variation of the seasonal component from 2 to -2, as observed in the time series analysis of temperature data, reflects the magnitude and direction of seasonal fluctuations relative to the long-term trend. The values from 2 to -2 represent the degree to which the seasonal component influences the temperature data. A value of 2 indicates that the seasonal effect increases the temperature by 2 degrees above the trend at a certain point or period in the cycle. Conversely, a value of -2 suggests a decrease in temperature by 2 degrees below the trend due to the seasonal effect. The range between these values (-2 to 2) illustrates the full amplitude of the seasonal variation, providing a clear picture of how much the temperature rises and falls over a typical cycle. Positive values in the seasonal component indicate periods when the season contributes to higher temperatures than the overall trend. This might correspond to the warmer months of the year, such as summer or a peak season specific to the region's climate. On the other hand, negative values suggest that the seasonal effect is lowering the temperatures below the trend, likely corresponding to cooler periods like winter. Since the seasonal component is analysed over an annual cycle, the variation from 2 to -2 also reflects the timing and duration of different seasons. The points where the seasonal component is highest (around 2) mark the peak of the warm season, while the lowest points (around -2) signify the coldest part of the year. The transition between these extremes shows the gradual temperature change as seasons progress. These seasonal variations are relative to the overall trend of the data.

Understanding the range and behaviour of the seasonal component is crucial for several reasons. It helps predict temperature patterns, plan for seasonal impacts, and understand how the typical seasonal cycle might change over time. For instance, shifts in the amplitude or timing of these seasonal fluctuations could indicate changes in climate patterns or the influence of urban development on local temperatures. The variation of the seasonal component from 2 to -2 in temperature data indicates the strength and direction of seasonal influences on temperature relative to the long-term trend. This analysis helps understand the cyclical nature of temperature changes, providing valuable insights for climate studies, urban planning, and environmental policy-making.

The residuals (Fig 1d) represent the irregularities or random variations in the temperature data that cannot be attributed to the trend or seasonal components. This plot is a key indicator of the volatility and unexpected fluctuations in the temperature data. These residuals are essentially the differences between the actual observed temperatures and the trend and seasonal components. The values from 10 to -10 indicate the extent of deviations from the expected pattern as defined by the trend and seasonal components. A residual of 10 means that the actual temperature was 10 degrees higher than the trend component for that specific point in time. Conversely, a residual of -10 indicates that the actual temperature was 10 degrees lower than the trend component. This range shows the maximum extent of these irregularities over the observed period.

Residuals are crucial for understanding the unpredictability inherent in the temperature data. They might include the effects of random or one-off events such as unusual weather patterns or sudden changes in local conditions (like urban development or deforestation). High positive values suggest periods of unexpected warming, while large negative values indicate unexpected cooling. Large positive or negative residuals could indicate extreme weather events or abrupt climate anomalies. For instance, an unseasonably hot day might result in a large positive residual, while an unexpectedly cold day might produce a significant negative residual.

The decomposition of the temperature time series into its constituent parts has revealed a clear trend, a definitive seasonal cycle, and random variations represented by the residuals. The trend component indicates the general direction of temperature changes over the period, the seasonal component highlights the regular pattern occurring annually, and the residuals show the erratic and unpredictable elements of the temperature data. Together, these results provide a comprehensive understanding of the temperature dynamics in the focus area of the study.

Stationarity check The Augmented Dickey-Fuller (ADF) test was conducted to determine the stationarity of the Visakhapatnam temperature time series. The results are critically important as they help understand whether the series has a unit root, a characteristic of a non-stationary series. The outcomes of the test are as follows:

ADF Statistic: The calculated value is -11.450497487592013. This statistic is a negative number, a primary indicator of stationarity. The more negative this statistic, the stronger the rejection of the hypothesis that there is a unit root at some confidence level.

p-value: The p-value obtained from the test is 5.874342587751549e-21. In hypothesis testing, the p-value helps determine the significance of the results. A low p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, suggesting it can be rejected.

The null hypothesis for the ADF test is that the time series has a unit root and is non-stationary. Given the extremely low p-value and the highly negative ADF statistic, we reject the null hypothesis. This suggests that the time series does not have a unit root and is stationary.

The stationarity of the time series implies that the statistical properties such as mean, variance, and autocorrelation are constant over time. This is crucial for building reliable predictive models, as many time series forecasting methods assume stationarity. Stationary data are easier to model and can be used to draw more reliable conclusions.

The result of the ADF test suggests that the temperature data is suitable for modelling without needing to be differenced to achieve stationarity. This expands the range of potential models that can be effectively applied for analysis and forecasting.

The analysis of the trend component in the Visakhapatnam temperature data reveals significant insights into the long-term behaviour of temperatures in the studied area. The following key statistics were derived from the trend component:

Mean Trend over the Period: The mean of the trend component is calculated to be 32.90. This value represents the average level around which the temperature varied over the entire study period. In the context of the temperature analysis for Visakhapatnam, this mean value indicates the area's general climatic conditions. A mean trend of 32.90 suggests a relatively high average temperature, which might be characteristic of a warm urban climate or a region experiencing higher

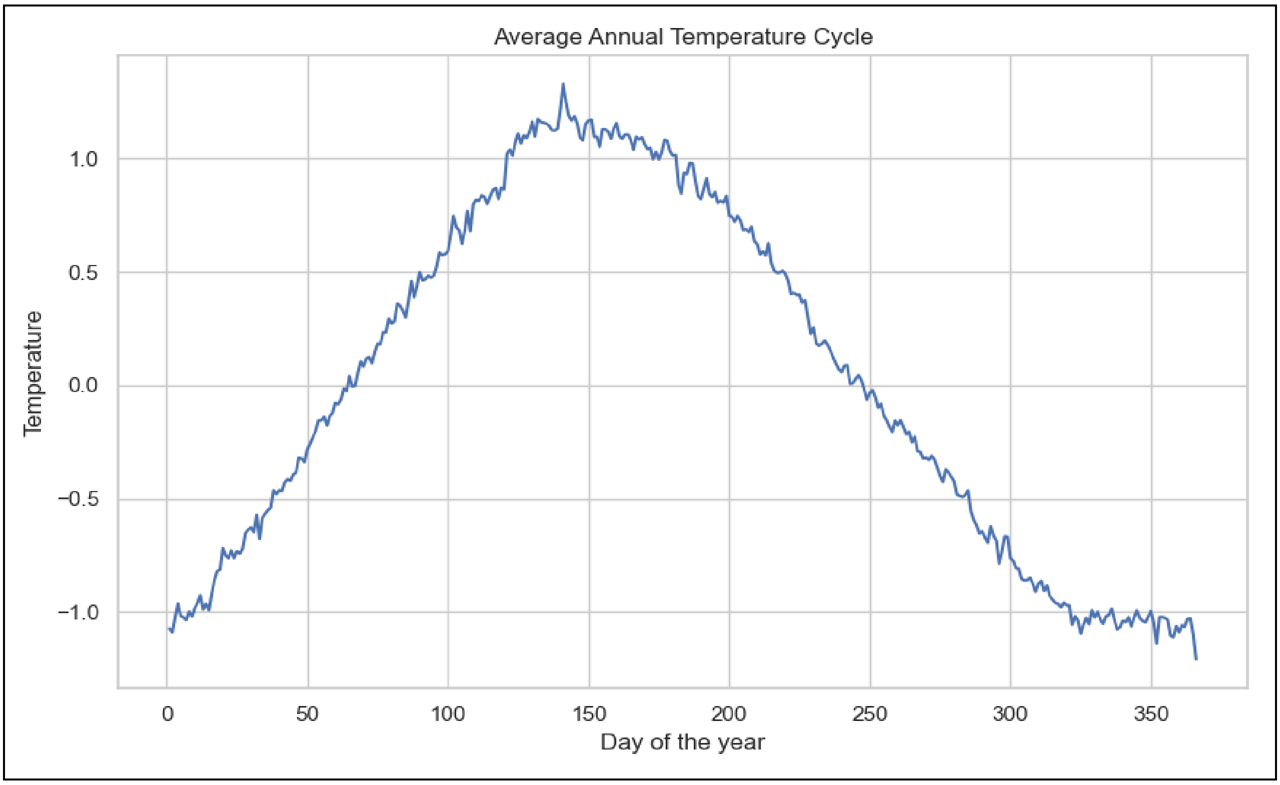

Figure 2: Average annual temperature cycle.

temperatures. This high average temperature may have significant implications for urban planning, public health, and energy consumption patterns in the area

Trend Variability (Standard Deviation): The standard deviation of the trend component is 0.58. This measure of variability indicates how much the temperature deviates from the mean trend over time. A standard deviation of 0.58 points to a relatively stable trend with minor fluctuations around the mean. This level of variability suggests that, despite seasonal and other short-term changes, the overall temperature in the area does not experience extreme variations over the long term.

The analysis of the trend component of the temperature data for Visakhapatnam Airport highlights a relatively high and stable temperature regime over the study period. This finding is essential for comprehending the long-term thermal environment of the urban area, informing policy decisions, and preparing for future climatic scenarios.

The examination of the seasonal component in the maximum temperature data of Visakhapatnam Airport provides a detailed understanding of the seasonal patterns and their impact on temperature variations. The key aspect of this analysis is the focus on the average annual temperature cycle derived from the seasonal component of the time series.

The analysis involved grouping the seasonal component by the day of the year and calculating the mean for each day. This process resulted in a plot representing the average annual temperature cycle. This visualization is crucial as it encapsulates the typical seasonal temperature variations experienced throughout the year. The annual average temperature cycle can be seen in Fig 2.

The plotted annual cycle clearly demonstrates how temperatures fluctuate over the course of a year. It highlights the periods of the year when temperatures are generally higher or lower, effectively capturing the essence of each season in terms of temperature behavior.

The plot in Fig 2 allows for identifying key points in the year – such as the hottest and coldest days on average. These points are crucial for understanding the extremities of seasonal temperature variations and their timing. The seasonal component analysis of urban temperature data offers a comprehensive view of how temperatures vary throughout the year. This analysis not only aids in understanding the inherent seasonal dynamics of the urban area's climate but also serves as a foundational element for further studies and applications in various sectors affected by temperature variations.

The analysis of residuals, derived from the decomposition of the maximum temperature data of Visakhapatnam airport, provides insight into the irregular or unexplained variations in the temperature time series. The residuals are the differences between the actual observed temperatures and the trend and seasonal components. The following summary statistics were obtained from the residuals analysis:

Count: The total number of residual data points analysed is 18,854. This high count indicates a substantial dataset, allowing for a robust analysis of the residuals.

Mean of Residuals: The mean value of the residuals is approximately 0.004587. This value, being close to zero, suggests that, on average, the decomposed trend and seasonal components explain the temperature variations. A mean close to zero in residuals is often indicative of effectiveness.

Standard Deviation: The standard deviation of the residuals is 2.617951. This measure indicates the typical deviation of the residual values from the mean. A standard deviation of about 2.62 suggests that fluctuations are still captured as residuals while much of the temperature variation is explained.

Minimum and Maximum Values: The minimum and maximum values of the residuals are -9.764476 and 11.858538, respectively. These values show the range of the residuals and indicate the largest underestimation and overestimation compared to the actual temperatures. Such extremes can be due to anomalous weather events.

Interquartile Range: The 25th percentile (Q1) and the 75th percentile (Q3) are -1.811446 and 1.738441, respectively. About 50% of the residuals fall within this range. This interquartile range measures the central tendency of the residuals and gives an idea of the typical magnitude of the deviations from the predicted trends.

The residuals analysis offers a deeper understanding of the aspects of temperature variation that are not explained by the trend and seasonal components. This analysis is crucial for evaluating the accuracy, understanding unexplained variations in the data, and guiding further improvements in the approach.

In the domain of time series analysis, elucidating autocorrelation and partial autocorrelation patterns, particularly in the residuals, is indispensable. These metrics offer profound insights into the latent structures and correlations within the residuals of urban temperature data, thereby informing the robustness and comprehensiveness of the analysis. Here, we delve into the analyses derived from the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the residuals.

The ACF plot is instrumental in assessing the correlation of the series with its own lagged values. By extending this analysis to the residuals, one can discern any systematic, non-random pattern. In this study, the ACF plot for the residuals was delineated for up to 50 lags (Fig 3). The decay pattern of autocorrelations, or their persistence, can be pivotal in determining whether the residuals represent white noise characterized by a lack of autocorrelation. Notably, pronounced autocorrelation at specific lags could suggest that the residuals harbour a systematic pattern, potentially implying that the current analysis might have overlooked certain components or underlying seasonal or cyclical dynamics.

The Autocorrelation Function (ACF) plot for the temperature data exhibits a gradual decline in correlation values as the lag increases, consistent with the nature of time series data. An ACF value of 1 at lag 0 is expected, as it represents the perfect correlation of the series with itself at the same time point, serving as the baseline for the ACF plot without providing additional meteorological insights. A significant ACF value at lag 1 indicates a strong correlation between the temperature on a given day and the previous day's temperature, suggesting a high degree of day-to-day persistence in weather conditions. This is characteristic of

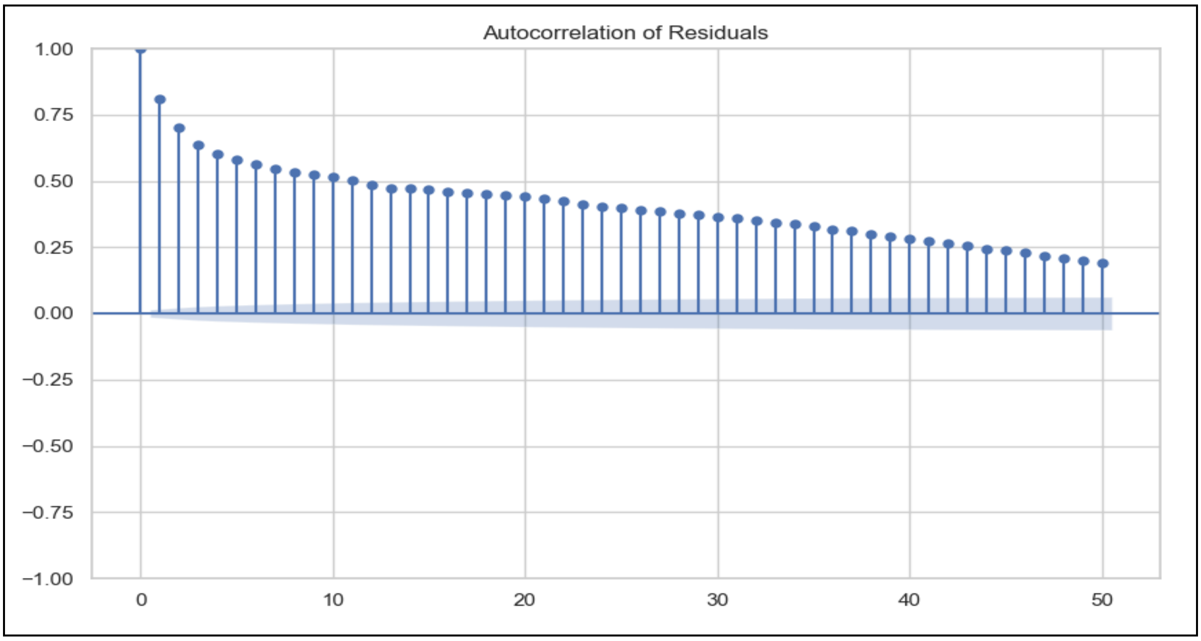

Figure 3: Autocorrelation of residuals.

stable atmospheric situations, where persistent high or low-pressure systems dominate, maintaining similar temperature patterns over consecutive days.

The relatively high but gradually decreasing ACF values up to lag 5 suggest that the influence of past temperatures extends to subsequent days, albeit with diminishing strength. This pattern likely reflects the typical lifecycle of weather systems, which often impact a region's climate over several days before dissipating or transitioning. The slow decay in correlation values indicates that the prevailing weather and temperature conditions exert a lingering influence over the course of approximately one week.

The presence of moderate autocorrelation at lag 10 implies that the temperature from 10 days prior continues to exert a discernible impact on current temperature conditions. This finding may indicate longer-lasting weather patterns or the residual effects of large-scale climate phenomena such as prolonged heatwaves, cold spells, or extended periods of stable high-pressure conditions.

The continued but gradually diminishing autocorrelation up to lag 40 suggests that past weather conditions exert a long-lasting influence, though with decreasing intensity. This may reflect the impact of broader climatic patterns, such as oceanic cycles or extended seasonal transitions, on the region's climate.

At lag 50, the ACF value is significantly lower but remains positive, indicating a weak yet persistent relationship between current temperatures and those from 50 days earlier. This suggests the presence of very persistent seasonal trends or the influence of slow-moving climatic phenomena.

The above ACF analysis provides valuable insights into the persistence and memory of temperature conditions from a meteorological perspective, highlighting both the short-term continuity and the long-term influence of past temperatures. This information is crucial for improving weather forecasting, understanding climate behaviour, and informing strategic planning for meteorological and climatic impacts.

The PACF plot, elucidating the partial correlation of the series with its lagged values while controlling for the values at shorter lags, is particularly telling. It helps isolate the correlation at each individual lag, unencumbered by shorter lag correlations.

The Partial Autocorrelation Function (PACF) plot of the residuals was constructed up to 50 lags, as illustrated in Fig 4. Significant peaks at specific lags within this plot suggest the potential existence of autoregressive components within the time series. The PACF is particularly valuable for identifying the appropriate order of autoregressive terms, which is essential if further modelling of the residuals is required.

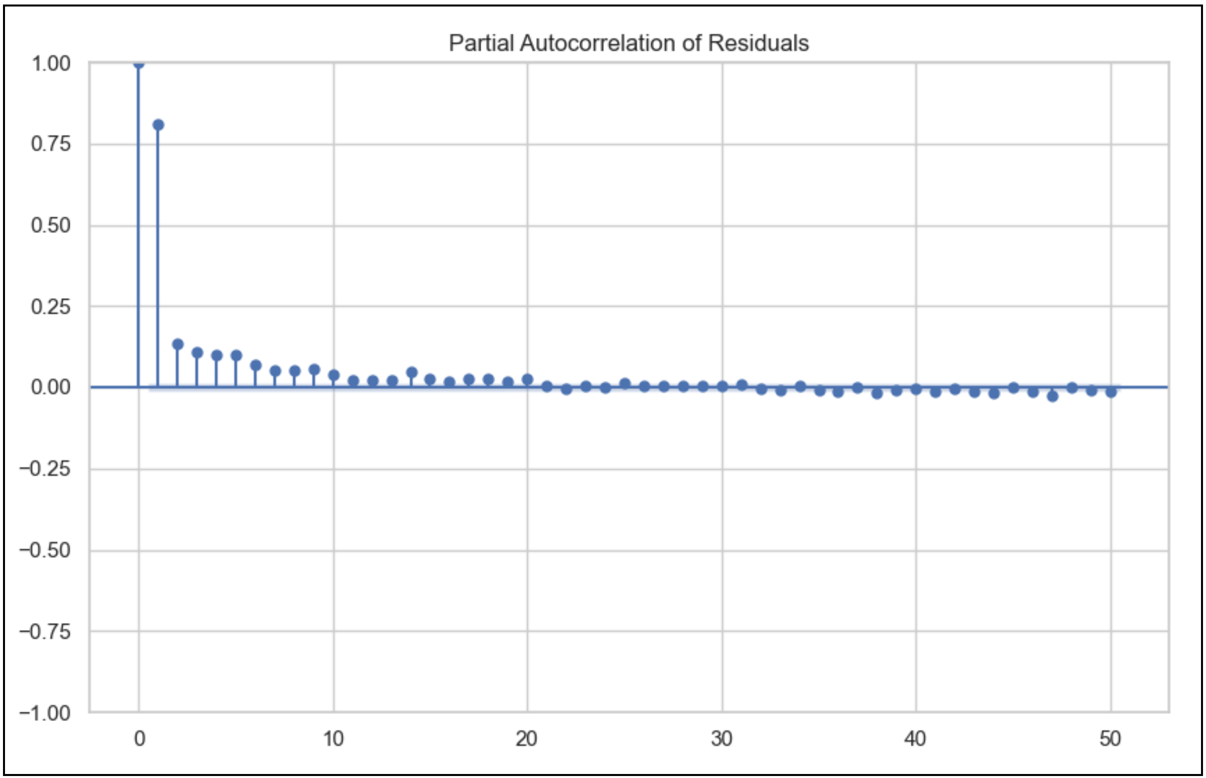

Figure 4: Partial Autocorrelation of residuals.

Analyzing the Partial Autocorrelation Function (PACF) from a meteorological standpoint involves interpreting the autoregressive nature of temperature data in the context of atmospheric and environmental factors. A high PACF value at lag 1 indicates a strong influence of the previous day's temperature on the current day's temperature, suggesting significant day-to-day continuity in weather conditions. Meteorologically, this pattern is often associated with persistent atmospheric systems, such as high-pressure zones, which stabilize weather conditions over several consecutive days.

The diminishing yet still observable PACF values at lags 2 and 3 imply that the influence of past temperatures extends beyond just the previous day, albeit with progressively decreasing strength. This pattern may reflect the gradual transition of weather systems or the residual effects of atmospheric conditions, such as a heatwave, that continue to impact temperatures for several days before the system fully dissipates or shifts.

The presence of a similar level of partial autocorrelation at lag 4, comparable to that observed at lags 2 and 3, could indicate the typical duration of certain weather patterns in the region. For example, some weather systems may typically exert their influence over four days before moving on or breaking down, affecting the region's temperature profile during this timeframe.

The presence of weak partial autocorrelation at these lags could point to the impact of longer-lasting meteorological phenomena. For example, extended periods of stable weather or prolonged climate patterns like heatwaves or cold spells might affect the temperatures consistently over a week.

The absence of partial autocorrelation from lag 11 onwards suggests that the temperature on a given day is not significantly influenced by temperatures from more than 10 days prior. This pattern aligns with the typical behaviour of weather systems, which generally do not directly influence local temperatures beyond this timeframe. This indicates that the impact of past weather conditions tends to dissipate within approximately 10 days, reflecting the transient nature of most atmospheric systems.

The PACF analysis in a meteorological context suggests a strong day-to-day continuity in temperature, likely influenced by persistent weather conditions, with diminishing influences over a

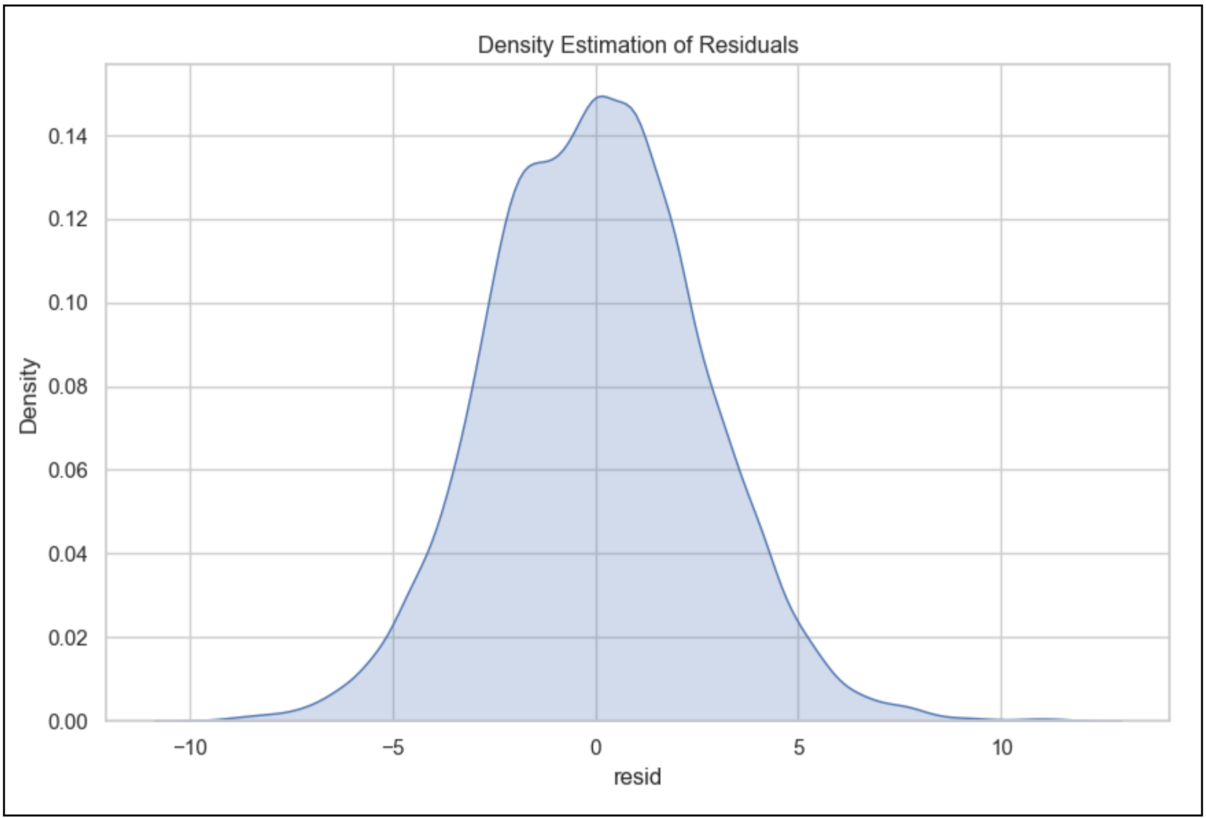

Figure 5: Density estimation of residuals.

period of up to 10 days. This understanding can be crucial for short-term weather forecasting and for studying the behaviour of local weather systems and their impact on temperature patterns.

In this research work, the residuals from the urban temperature time series analysis is understood by employing a Kernel Density Estimation (KDE) plot. The KDE plot is a non-parametric way to estimate the probability density function of a random variable, in this case, the residuals of maximum temperature data. The KDE plot (Fig. 5) offers a smoothed view of the residuals' distribution, highlighting the residuals' concentration around zero, suggesting that the model is generally accurate. Unlike a histogram, which is also used for understanding distributions but can be sensitive to bin sizes, the KDE plot offers a more refined and interpretable visualization of how the residuals are distributed around the mean.

The shape and spread of the KDE plot provide insights into the nature of the residuals. For instance, a narrow, sharp peak around zero would indicate that most residuals are small, suggesting that the model predictions are generally close to the actual values. On the other hand, a wider spread or multiple peaks might suggest more variability in the model's performance or the presence of different regimes or behaviours in the temperature data that couldn’t be captured uniformly.

In a meteorological context, the residuals represent the unpredictable or unexplained variations in temperature after accounting for the regular pattern (trend and seasonality). These might be due to random weather fluctuations, unique climatic events, or other non-systematic factors. The tails of the distribution in the KDE plot can help identify outliers or extreme values in the residuals. These represent days when the actual temperature was significantly different from the trends. Investigating these anomalies could lead to insights into rare but impactful meteorological events, such as heatwaves or storms.

The bell-shaped distribution of the residuals, tapering to zero around ±9, suggests that most temperature deviations are within this range. This indicates a relatively high level of accuracy, as extreme deviations (residuals) are uncommon. Meteorologically, this implies that while the temperature patterns (trend and seasonality) are captured, there are inherent fluctuations in daily temperatures that cannot be perfectly predicted, possibly due to random or unforeseen weather events.

The peak density at around 0.15, with most residuals clustering around zero, indicates that the trends are generally close to the actual temperatures. This is a desirable feature in any trend analysis, particularly in meteorological forecasting, where precision is critical. A concentration of values near zero in the residuals suggests that the unexplained variation is mostly random noise, typical in weather data, due to its chaotic nature and the influence of numerous uncontrollable and unpredictable factors.

The noted depression on the left side of the curve, between 0 to -2 residuals, is an interesting anomaly. It indicates a slightly lower frequency of mild negative residuals than a perfectly symmetrical distribution suggests. In meteorological terms, this could imply that there are fewer instances where a slight overestimation of the temperature is captured compared to slight underestimations. This asymmetry might reflect certain climatic conditions or environmental factors that are not fully captured, leading to a marginally higher occurrence of days where the actual temperature is slightly cooler than predicted.

Fig 6 shows the boxplot of the residuals for the maximum temperature dataset used in this study of Visakhapatnam Airport. A median value close to zero in the boxplot indicates the overall accuracy of the temperature trends. In meteorological terms, this suggests that, on average, the forecast trends are well-aligned with the actual temperatures. The median being at or near zero strongly indicates the absence of any bias.

The box representing the IQR, extending from around -2 to +2, signifies that 50% of the residuals fall within this range. This relatively narrow spread indicates that most of the model's temperature predictions are within ±2 degrees of the actual temperatures. In the context of weather forecasting and climate modelling, this level of accuracy is generally considered good, especially for daily temperature predictions in urban areas.

The description of the box as a perfect rectangle suggests a symmetrical distribution of residuals. The absence of skewness in the distribution is desirable as it suggests no consistent directional bias.

From a forecasting perspective, the trend performance, as the boxplot indicates, is quite reliable for typical weather conditions. However, it's important to note that the IQR does not capture the behaviour of the trends under more extreme weather conditions, which outliers may represent.

The upper and lower whiskers of the boxplot, extending to 7 and -7, indicate the range within which the bulk of the temperature residuals lie, excluding outliers. In meteorological terms, this means that the trends are usually within 7 degrees of the actual temperature. This range might be considered acceptable for general forecasting but suggests that significant deviations occur more frequently than ideal, possibly during unusual or extreme weather conditions.

The circles above and below the whiskers represent outliers, which are residuals that lie beyond the expected range of variability based on the IQR.

• Outliers above the upper whisker (up to 12.5) indicate days when the actual temperature was significantly higher than the trend. These could correspond to unexpected heatwaves or other anomalous warm weather events not captured by the trends.

• Outliers below the lower whisker (down to -10) represent days when the actual temperature was much lower than the trend, possibly due to unforeseen cold days, storms, or other unusual cooling phenomena.

The presence of outliers is particularly important in meteorology as they often correspond to extreme weather events that can have significant implications for public safety, energy demand, and general preparedness. The range and frequency of these outliers could also indicate the volatility of the region's climate or the presence of

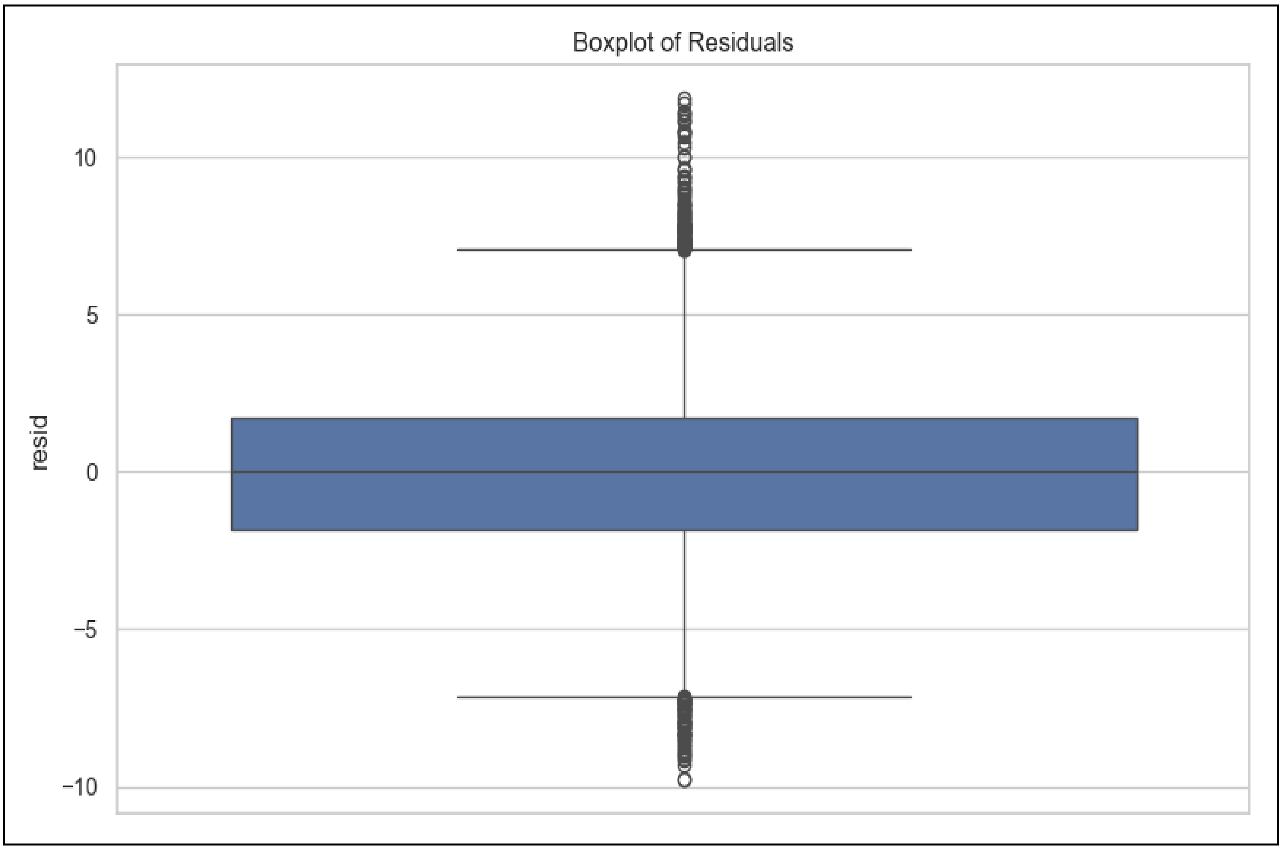

Figure 6: Boxplot of residuals.

microclimates, which might be influenced by geographical features, urbanization, or changing climate patterns. The outliers, particularly those representing extreme temperature deviations, are crucial for understanding the area's risk of extreme weather events. They help meteorologists and urban planners identify potential vulnerabilities and prepare more effectively for unusual weather.

The boxplot of residuals from the temperature data of Visakhapatnam, with its detailed features, including the position of the whiskers and the identification of outliers, provides a nuanced understanding of the trend performance and the nature of temperature variability in the area. Analyzing these features from a meteorological perspective is vital for assessing the reliability of temperature predictions, understanding the occurrence of extreme weather events, and guiding improvements in forecasting models and strategies.

4. Conclusion

This study provides a comprehensive analysis of the maximum temperature data from Visakhapatnam using advanced time series methods, offering significant contributions to our understanding of urban climatology in a tropical coastal city. The meticulous decomposition of temperature data into trend, seasonal, and residual components has yielded valuable insights into the underlying patterns and dynamics of urban temperatures over an extended period.

This research's key findings include identifying long-term temperature trends, where the trend analysis revealed a persistent upward trajectory in maximum temperatures, indicating a gradual warming trend in Visakhapatnam. This finding is crucial as it underscores the impact of global climate change and urbanization on local temperature patterns. The relatively high and stable average temperature identified in the trend component suggests a warming climate, which could exacerbate the urban heat island effect and lead to more frequent and severe heat-related health risks. This insight is essential for urban planners and policymakers as they develop strategies to mitigate the impacts of rising temperatures and enhance urban resilience. Additionally, the detailed understanding of seasonal variations provided by the seasonal analysis highlighted the periods of the year most susceptible to extreme temperatures. The predictable nature of these seasonal patterns allows for better planning and preparedness in agriculture, energy management, and public health sectors. For instance, understanding the timing and intensity of seasonal temperature peaks can help optimize energy consumption for cooling and prepare for potential heat waves.

Furthermore, the assessment of residual variability identified the irregular and unpredictable components of temperature variations, which are not captured by the trend or seasonal models. These residuals often reflect sudden weather events or local climatic anomalies, which are critical for improving the accuracy of temperature forecasts and understanding the full range of temperature variability. Significant outliers in the residuals emphasise the need for further research into extreme weather events and their underlying causes.

List of Recommendations

(a) Enhance Urban Green Spaces: Increasing green spaces within urban areas is recommended based on the observed trend of rising temperatures and the significant impact of seasonal variations. Urban parks, green roofs, and street trees can mitigate the urban heat island effect by providing natural cooling and reducing heat absorption by built structures. This strategy is particularly important for areas most affected by extreme temperature variations.

(b) Implement Reflective and Green Roofing Systems: The study's findings on the relatively high and stable temperature trend over time suggest a need for measures that can reduce heat accumulation in buildings. Promoting reflective roofing materials and green roofs can lower building surface temperatures, reduce indoor cooling demands, and contribute to overall urban cooling. Incentives for adopting these technologies should be considered in urban planning policies.

(c) Develop Heat-Resilient Urban Infrastructure: Identifying outliers representing extreme weather events highlights the necessity of heat-resilient infrastructure. Urban planners should prioritize the development of infrastructure that can withstand extreme temperatures, including materials that resist heat deformation and expanding shaded public spaces to protect citizens during heatwaves.

(d) Strengthen Public Health Preparedness: Given the observed irregularities in temperature data, which may indicate unexpected and extreme weather events, it is crucial to strengthen public health preparedness. Establishing cooling centres, enhancing early warning systems for heatwaves, and conducting public awareness campaigns about the dangers of extreme heat are essential measures to protect vulnerable populations.

(e) Integrate Climate-Responsive Urban Planning: The persistence of temperature trends over long periods, as revealed by the autocorrelation and partial autocorrelation analyses, suggests the need for long-term urban planning strategies considering ongoing climate changes. Policies should incorporate climate resilience by promoting energy-efficient buildings, designing urban layouts that facilitate airflow, and reducing the urban heat island effect through reflective surfaces and vegetation.

By integrating these strategies, the city can improve its resilience to heat-related risks and create a more sustainable and livable urban environment.

The findings pave the way for further research in this domain, aiming to improve the prediction of urban weather patterns and thereby contribute to developing more resilient and sustainable urban environments.

References

Oke, T. R. (1982). The energetic basis of the urban heat island. Quarterly Journal of the Royal Meteorological Society, 108(455), 1-24.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. Bulletin of the American Meteorological Society, 93(12), 1879-1900.

Mills, G. (2014). Synoptic climatology and the analysis of weather data. In Urban Climatology: A Review of Methods and Topics. Cambridge University Press.

Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. Remote Sensing of Environment, 86(3), 370-384.

Rupp, D. E., Li, S., Mote, P. W., Massey, N., Sparrow, S. N., & Wallom, D. C. H. (2017). Influence of the urban heat island effect on regional climate projections: A case study for Greater London, United Kingdom. Journal of Climate, 30(10), 3921-3935.

Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305(5686), 994-997.

Kotharkar, R., & Bagade, A. (2018). Urban heat island studies in South Asia: A critical review. Urban Climate, 24, 1011-1026.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Chatfield, C. (2016). The Analysis of Time Series: An Introduction. CRC Press.

Mills, G. (2008). Luke Howard and the climate of London. Weather, 63(6), 153-157.

Arnfield, A. J. (2003). Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. International Journal of Climatology, 23(1), 1-26.

Eliasson, I. (2000). The use of climate knowledge in urban planning. Landscape and Urban Planning, 48(1-2), 31-44.

Ren, C., Ng, E., & Katzschner, L. (2011). Urban climatic map studies: A review. International Journal of Climatology, 31(15), 2213-2233.

Santamouris, M. (2014). Cooling the cities – A review of reflective and green roof mitigation technologies to fight heat island and improve comfort in urban environments. Solar Energy, 103, 682-703.

Stone, B., Hess, J. J., & Frumkin, H. (2010). Urban form and extreme heat events: Are sprawling cities more vulnerable to climate change than compact cities? Environmental Health Perspectives, 118(10), 1425-1428.

Tan, J., Zheng, Y., Tang, X., Guo, C., Li, L., Song, G., ... & Chen, H. (2010). The urban heat island and its impact on heat waves and human health in Shanghai. International Journal of Biometeorology, 54(1), 75-84.

Wilby, R. L. (2003). Past and projected trends in London's urban heat island. Weather, 58(7), 251-260.

Zhou, B., Rybski, D., & Kropp, J. P. (2017). The role of city size and urban form in the surface urban heat island. Scientific Reports, 7(1), 1-9.

Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press.

Landsberg, H. E. (1981). The urban climate. International Geophysics Series, 28.

Grimmond, S. (2007). Urbanization and global environmental change: Local effects of urban warming. Geographical Research, 45(1), 83-88.

Li, D., & Bou-Zeid, E. (2013). Synergistic interactions between urban heat islands and heat waves: The impact in cities is larger than the sum of its parts. Journal of Applied Meteorology and Climatology, 52(9), 2051-2064.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. 2nd ed. OTexts.

Akbari, H., Pomerantz, M., & Taha, H. (2001). Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas. Solar Energy, 70(3), 295-310.

Sailor, D. J. (2014). A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. International Journal of Climatology, 34(6), 1501-1517.

Gartland, L. (2010). Heat Islands: Understanding and Mitigating Heat in Urban Areas. Earthscan.

Middel, A., & Krayenhoff, E. S. (2019). Micrometeorological determinants of pedestrian thermal exposure during record-breaking heat in

Tempe, Arizona: Introducing the MaRTy observational platform. Science of The Total Environment, 687, 137-151.

Rizwan, A. M., Dennis, L. Y. C., & Liu, C. (2008). A review on the generation, determination and mitigation of urban heat island. Journal of Environmental Sciences, 20(1), 120-128.

Yow, D. M. (2007). Urban heat islands: Observations, impacts, and adaptation. Geography Compass, 1(6), 1221-1256.

Zhao, L., Lee, X., Smith, R. B., & Oleson, K. (2014). Strong contributions of local background climate to urban heat islands. Nature, 511(7508), 216-219.