

Article

Not peer-reviewed version

Forecasting Temperature Trends Using SARIMAX: A Case Study in Ahmedabad City, India

[Vyom Shah](#) , [Nishil Patel](#) , [Dhruvin Apurva Shah](#) , [Debabrata Swain](#) ^{*} , Manorama Mohanty ,
[Biswaranjan Acharya](#) , [Vassilis C. Gerogiannis](#) ^{*} , [Andreas Kanavos](#)

Posted Date: 18 July 2024

doi: 10.20944/preprints2024071521.v1

Keywords: temperature forecasting; weather forecasting; time series; augmented Dickey-Fuller test; seasonal autoregressive integrated moving average with exogenous factors (SARIMAX); root mean squared error; seasonality; climate change



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Forecasting Temperature Trends Using SARIMAX: A Case Study in Ahmedabad City, India

Vyom Shah ¹, Nishil Patel ¹, Dhruvin Shah ¹, Debabrata Swain ^{1,*}, Manorama Mohanty ², Biswaranjan Acharya ³, Vassilis C. Gerogiannis ^{4,*} and Andreas Kanavos ⁵

- ¹ Computer Science and Engineering Department, Pandit Deendayal Energy University, Gandhinagar 382007, India; vyom.sce21@sot.pdpu.ac.in (V.S.); nishil.pce21@sot.pdpu.ac.in (N.P.); dhruvin.sce21@sot.pdpu.ac.in (D.S.)
- ² Indian Metrological Department Bhubaneswar 751020, India; m.mohanty@imd.gov.in
- ³ Department of Computer Engineering-AI, Marwadi University, Rajkot 360003, India; biswaranjan.acharya@marwadieducation.edu.in
- ⁴ Department of Digital Systems, University of Thessaly, Larissa, Greece
- ⁵ Department of Informatics, Ionian University, Corfu, Greece; akanavos@ionio.gr
- * Correspondence: debabrata.swain@sot.pdpu.ac.in (D.S.); vgerogian@uth.gr (V.C.G.)

Abstract: Globalization and industrialization have significantly disturbed the environmental ecosystem, leading to critical challenges such as global warming, extreme weather events, and water scarcity. Forecasting temperature trends is crucial for enhancing the resilience and quality of life in smart sustainable cities, enabling informed decision-making and proactive urban planning. This research specifically targets Ahmedabad city in India and employs the Seasonal Auto Regressive Integrated Moving Average with eXogenous factors (SARIMAX) model to forecast temperatures over a ten-year horizon using two decades of real-time temperature data. The stationarity of the dataset was confirmed using the Augmented Dickey-Fuller test, and the Akaike Information Criterion (AIC) method helped identify the optimal seasonal parameters of the model, ensuring a balance between fidelity and predictive accuracy. The model achieved an RMSE of 1.0265, indicating high accuracy within the typical range for urban temperature forecasting. This robust measure of error underscores the model’s precision in predicting temperature deviations, which is particularly relevant for urban planning and environmental management. The findings provide city planners and policymakers with valuable insights and tools for preempting adverse environmental impacts, marking a significant step towards operational efficiency and enhanced governance in future smart urban ecosystems. Future work may extend the model’s applicability to broader geographical areas and incorporate additional environmental variables to refine predictive accuracy further.

Keywords: temperature forecasting; weather forecasting; time series; augmented Dickey-Fuller test; Seasonal AutoRegressive Integrated Moving Average with eXogenous factors (SARIMAX); root mean squared error; seasonality; climate change

1. Introduction

The phenomenon of climate change, characterized primarily by a significant rise in global temperatures, presents unprecedented challenges to our planet. This escalation in temperature is especially perilous in densely populated and industrial areas, where the effects of heatwaves are exacerbated by the emission of greenhouse gases. Such environmental shifts are leading to severe, multifaceted consequences. According to the National Oceanic and Atmospheric Administration (NOAA), the period from 2018 to 2020 witnessed over 3,000 death toll in the United States attributable to heat-related complications [1]. Currently, urban areas, which house more than 55% of the global population — a figure projected to swell to 68% by 2050 — are on the frontline, facing increased risks of premature mortality and heat-induced illnesses [2].

The primary driver of global warming due to human-induced greenhouse gas emissions has resulted in a 1.1°C increase in the average global temperature over the decade spanning 2011 to 2020 [3]. This warming trend not only raises air and ocean temperatures but also signals severe implications for urban governance and infrastructure. This study is motivated by the urgent need for government agencies to grasp the potential consequences of rising temperatures fully. By anticipating these changes,

local governments can implement strategic modifications to urban infrastructure, enhancing resilience, reducing environmental impact, and safeguarding the well-being of city dwellers in an era of escalating global temperatures.

Moreover, the repercussions of increased temperatures extend beyond immediate health risks. Elevated temperatures and humidity exacerbate conditions such as eczema, psoriasis, and other dermatological disorders due to increased perspiration. The availability of comprehensive epidemiological data is crucial for the development of effective illness prevention strategies and treatment protocols for affected individuals. In parallel, the environmental toll of global warming manifests in the diminished capacity of plants to absorb and utilize moisture, leading to soil depletion, droughts, and the alarming acceleration of desertification. The process strips fertile land of its flora and fauna, converting it into barren deserts. India's alarming 122% increase in land loss to forest fires within a mere five-year span illustrates the devastating synergy between high temperatures, dry conditions, and the propensity for wildfires, which in turn lead to widespread deforestation [4].

The detrimental impacts of climate change do not halt at land degradation. Rising sea levels pose a direct threat to coastal communities, where approximately 40% of the global population resides. The melting of polar ice caps, coupled with the thermal expansion of ocean waters, contributes to a higher incidence of coastal erosion, elevated storm surge levels, and the enhanced severity of coastal storms. These phenomena jeopardize not only the physical safety of these regions but also their water security and economic stability [5]. Informed by the assessments of the Intergovernmental Panel on Climate Change (IPCC), governments worldwide are grappling with the task of understanding and mitigating the multifarious impacts of climate change [3].

The pivotal role of temperature projections in addressing the myriad challenges posed by climate change cannot be overstated [6]. The urgency of achieving Sustainable Development Goal 11 - SDG11 (i.e., "make cities and human settlements inclusive, safe, resilient, and sustainable") [7] underscores the importance of leveraging innovative technologies and methodologies for sustainable urban development. Accurate temperature projections emerge as a pivotal tool in this context, enabling cities to adapt to climate change proactively, improve urban planning and ensure the well-being of their inhabitants. By exploring limitations and potentials of existing forecasting models, the current research study seeks to bridge gaps in our understanding and application of climate predictive technologies. The goal is to not only enhance the accuracy of temperature predictions but also to contextualize these forecasts within the broader spectrum of climate change impact in urban areas. This endeavor underscores the critical need for interdisciplinary approaches in addressing the complex challenges posed by global warming, paving the way for innovative solutions that can protect and enhance the quality of life in urban settings across the globe [8].

More specifically, an advanced statistical forecasting model, such as the Seasonal AutoRegressive Integrated Moving Average with exogenous factors (SARIMAX), can offer promising avenues for enhancing weather forecasting accuracy. The SARIMAX model extends the capabilities of the traditional ARIMA model by incorporating both seasonal adjustments and exogenous variables. This inclusion allows the model to account for seasonal variations, which are particularly significant in meteorological data, and to incorporate external factors that influence weather patterns, such as environmental indices or economic indicators.

The integration of exogenous variables into the SARIMAX model enables it to capture the impact of events or inputs outside the typical meteorological datasets, providing a more holistic view of the factors that affect weather conditions. For instance, in urban settings like Ahmedabad, where rapid urbanization and environmental changes play a crucial role, SARIMAX can utilize data on urban heat islands, pollution levels, or land-use changes as exogenous inputs. This enhances the model's accuracy in predicting temperature fluctuations and other weather-related phenomena.

Furthermore, the seasonal component of the SARIMAX model ensures that the forecasting system recognizes and adapts to the inherent seasonal patterns in weather data, such as the intense heat during dry seasons or the cooling during monsoons. By accurately modeling these patterns, SARIMAX

can provide more reliable and precise forecasts, which are crucial for effective planning and decision-making in urban management and agricultural planning.

Implementing SARIMAX within our study involved utilizing the Python Statsmodels library, which provides robust tools for time-series analysis. This choice was driven by the library's comprehensive support for seasonal and exogenous modeling, making it an ideal choice for our complex forecasting needs.

The present research leverages a decade's worth of temperature data from Ahmedabad city in India, employing the SARIMAX model to forecast future temperature trends. Such time-series analysis-based models stand at the forefront of our efforts to predict and mitigate the adverse effects of climate change, enabling city planners and policymakers to devise informed, strategic responses to this global crisis [9]. The choice of Ahmedabad as a case study offers a unique lens through which to examine the implications of temperature forecasting in rapidly urbanizing regions. As one of India's most populous cities, Ahmedabad embodies the challenges and opportunities inherent in managing urban growth and environmental sustainability in the face of climate change. To sum up, the current research aims to contribute to the burgeoning field of urban climate studies, providing insights that may inform the development of more resilient, adaptive urban infrastructures capable of withstanding the current and future climatic shifts.

The remainder of the paper is structured to methodically unfold the research carried out, starting with Section 2 that offers a comprehensive review of existing studies, comparing various forecasting models and methodologies that have been previously employed to predict temperature changes and assess their impacts. Following this, Section 3 presents the proposed methodology, detailing the SARIMAX model's development, the rationale behind its selection, and the specific steps taken to tailor it for temperature forecasting in Ahmedabad. Section 4 evaluates the model's performance through rigorous testing against historical temperature data, employing statistical measures to assess its accuracy and reliability. Finally, Section 5 concludes the paper, summarizing the key findings and contributions of this study while also highlighting potential directions for future research to further refine and expand upon the predictive capabilities of temperature forecasting models.

2. Related Work

The advancements in forecasting methodologies have significantly enriched the arsenal available for tackling the intricacies of climate variability and weather prediction. This diversification reflects a growing recognition of the multifaceted nature of weather phenomena and the critical role of accurate predictions in mitigating their impacts. The exploration of innovative forecasting models, ranging from statistical analyses to cutting-edge computational techniques, illustrates the dynamic evolution of the field [10].

An innovative approach that combined seasonal auto-regression with wavelet decomposition on historical temperature data from Delhi highlighted the effectiveness of melding classical statistical tests, such as the Augmented Dicky Fuller Test, with advanced data decomposition techniques [11]. This methodology not only bolstered forecast accuracy but also set a precedent for the application of sophisticated models in climatic analysis. Concurrently, the exploration of wind speed forecasting models revealed that multivariate configurations offer superior performance over univariate models, advocating for the inclusion of multiple meteorological variables to refine predictions [12].

Neural networks, particularly Bi-LSTM (Bidirectional Long Short-Term Memory) models, have marked a significant breakthrough in weather forecasting, demonstrating that model training over varied temporal spans can drastically influence prediction accuracy [13]. The distinct advantage of shorter prediction intervals in improving air temperature forecasts highlights the critical role of data granularity and model architecture in predictive performance [14]. The success of hybrid models combining Genetic Algorithms with LSTM networks for rainfall prediction further validates the potential of integrating machine learning techniques with evolutionary algorithms to capture complex temporal trends, offering a superior alternative to conventional models [15].

The exploration of hybrid forecasting models reveals a strategic blend of methodologies to tackle the multifaceted nature of weather prediction. The application of classifier approaches, such as Chi square and Naive Bayes algorithms, illuminates the power of statistical learning for identifying intricate patterns within historical weather data, enhancing the precision of future forecasts [16]. This narrative is complemented by studies comparing the efficacy of statistical, artificial intelligence, and hybrid models, which collectively emphasize the importance of model selection tailored to specific forecasting challenges and data characteristics [17].

Advanced statistical methods have also been applied to weather forecasting, with certain studies pointing out that semi-average methods are well-suited for scenarios involving interval or imprecise data, providing a sturdy framework for trend analysis in neutrosophic statistics [18]. Similarly, the employment of the Markov-Chain Monte-Carlo approach alongside the Seasonal Autoregressive Integrated Moving Average model for wind speed estimation underscored the utility of Seasonal Auto-Regression in achieving precise high-speed predictions. This approach also highlighted the efficacy of Markov chain Monte-Carlo for short-term forecasting accuracy [19].

Comparative analyses of forecasting models serve to benchmark the performance and applicability of various advanced statistical approaches, including SARIMAX and GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) models. The GARCH model is particularly renowned for its ability to model financial time series data where volatility clustering—a phenomenon where high-volatility events are followed by high-volatility events and low-volatility events are followed by low-volatility events—is observed. This makes it highly suitable for risk management and option pricing in financial markets, contrasting with SARIMAX's utility in handling seasonal variations and external factors in meteorological and environmental data. These studies not only highlight the significance of incorporating exogenous factors in enhancing model robustness but also underscore the relevance of these methodologies in addressing real-world problems, such as urban planning and climate change adaptation [20]. The demonstrated correlation between forecasted and observed air temperatures in studies focused on specific locales further attests to the practical effectiveness of these models, validating their utility in operational settings [21].

The collective insights from studies such as the above mentioned underscore a dynamic and interdisciplinary pursuit of accuracy and reliability in weather forecasting. As the field gravitates towards integrating diverse models and methodologies, the emphasis on adaptability, precision, and practical application becomes increasingly evident. To conclude, the existing related work lays a foundational understanding that informs and motivates the present study, aiming to harness the strengths of the SARIMAX modeling for refined temperature prediction in the context of the Ahmedabad city in India, thereby contributing to the global research efforts in the fields of climate resilience and sustainable urban development.

3. Proposed Methodology

3.1. Data Selection

The cornerstone of our predictive analysis is anchored in a curated dataset procured from the India Meteorological Department (IMD), a repository acclaimed for its exhaustive and precise records of weather and climate data. This dataset encompasses monthly maximum temperature observations for Ahmedabad, Gujarat, spanning an extensive period from January 1993 to December 2022. This selection is not arbitrary; it encapsulates a period marked by significant climatic shifts, offering a rich canvas to explore temperature trends and anomalies. The dataset is further characterized by descriptive attributes—index, year, month, and maximum temperature—comprising a comprehensive collection of 360 records across four distinct features, as elaborated in Table 1.

Table 1. Features of the Temperature Dataset.

Feature	Description
INDEX	City Identifier
YEAR	Year of Record
MN	Month
MAX	Maximum Temperature (°C)

The dataset’s structure is designed to facilitate a granular analysis of temperature patterns. The index feature, identifying Ahmedabad, ensures geographical specificity, crucial for localized climate studies. The year and month attributes provide a temporal framework, enabling a detailed examination of seasonal dynamics and long-term climatic trends. Central to our study is the MAX feature, denoting the highest temperature recorded each month, which forms the basis of our forecasting endeavor.

This strategic selection of data grounds our research in a context that is both geographically and temporally pertinent. It ensures that our forecasting model is informed by a dataset that is not only comprehensive but also reflective of the nuanced climate dynamics specific to Ahmedabad. The period covered by the dataset, spanning three decades, is particularly significant, having a realistic possibility to capture a range of climatic phenomena that could influence temperature trends, from seasonal variations to more pronounced effects of global climate change.

By anchoring our analysis in this robust dataset, we aim to enhance the precision and relevance of our forecasting model. The geographical specificity to Ahmedabad, coupled with the dataset’s extensive temporal span, provides a solid foundation for identifying and forecasting temperature trends. It is this specific approach to data selection that underpins the reliability and applicability of our predictive analysis, so that the insights derived are accurate enough to be actionable for addressing the challenges posed by climate variability in the region.

3.2. *Data Preprocessing*

The essence of this research lies in accurately forecasting the maximum daily temperature for the Ahmedabad city in India. A meticulous data preprocessing phase was crucial to achieving this goal, detailed as follows:

3.2.1. *Data Cleaning and Transformation*

Initially, the dataset underwent a thorough cleaning process to better align with our forecasting model’s requirements. The ‘Index’ column, serving only as a placeholder for "Ahmedabad," was deemed redundant and subsequently removed. This step was followed by quality assurance checks to verify the consistency and integrity of temperature records, including outlier detection and correction to ensure data reliability.

Recognizing the periodic nature of our forecasting, a pivotal transformation involved the creation of a ‘Date’ column by merging ‘Year’ and ‘Month’ into a date-time format, which was then designated as the index. This restructured the dataset into a two-dimensional array, comprising ‘Date’ and ‘Max Temperature (MMAX)’ columns. The restructuring facilitated a more organized data analysis, directly supporting the statistical summary provided in Table 2, which outlines the descriptive statistics—mean, standard deviation, minimum, and maximum values—of the max temperature observed. Such a format simplifies subsequent analyses and enhances the clarity of our forecasting process. Although normalization or standardization processes were evaluated to facilitate model training, they were not applied in this instance due to the model’s robustness to scale variations. These practices would typically be considered where data scale might unduly influence learning processes.

Table 2. Descriptive Statistics of the Max Temperature.

	Mean	Standard Deviation	Min	Max
Max Temperature	34.42	4.18	26.10	43.80

In this study, the focus was placed on the maximum temperatures due to their critical impact on peak energy demand, heat-related health risks, and urban heat island effects, which are particularly significant in rapidly urbanizing regions like Ahmedabad. However, minimum temperatures also hold substantial meteorological and climatic importance, especially in areas experiencing significant urbanization, as they affect aspects like frost occurrences and heating needs. While this study concentrates on maximum temperatures to address specific urban planning and public health objectives, the inclusion of minimum temperatures could provide a more comprehensive view of the thermal environment and its broader implications.

3.2.2. Observing the Trend

An in-depth trend analysis, facilitated by a violin plot for one year, revealed a cyclical temperature pattern, depicted in Figure 1. To further dissect these patterns, a decomposition analysis was conducted, separating the time series into trend, seasonal, and residual components. This analysis, alongside statistical tests like the Mann-Kendall trend test, quantified the significance of observed trends, confirming their persistence and statistical relevance over time.

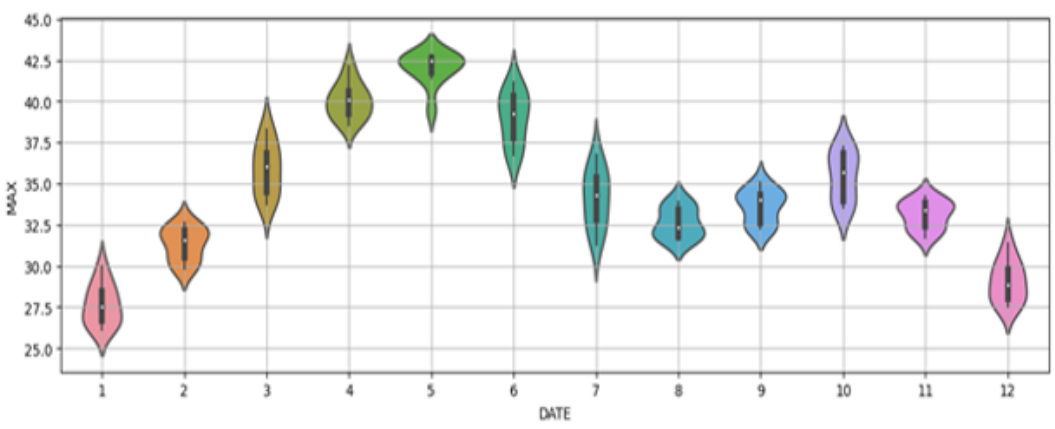


Figure 1. Violin Plot for 1 Year.

3.2.3. Stationarity Check Using Augmented Dickey Fuller Test

Ensuring data stationarity is a cornerstone for the application of SARIMAX in time series forecasting. Stationary data allow the model to effectively identify and leverage meaningful patterns, relationships, and trends inherent in the dataset. Achieving stationarity—through differentiation and transformation of the data—stabilizes the model parameters, thereby bolstering the reliability and accuracy of its predictions. This foundational aspect of SARIMAX underscores the necessity of stationarity for the model’s operational efficacy and the trustworthiness of its forecasts. To ascertain the stationarity of our time series data, we employed the Augmented Dickey-Fuller (ADF) test, a robust method for stationarity verification [22].

The ADF test, an enhanced iteration of the original Dickey Fuller test, is designed to assess whether a given time series exhibits stationarity. This test is imperative for our analysis; a non-stationary dataset would undermine the forecasting capability of the model, as SARIMAX relies on stationary data to generate forecasts. The ADF test scrutinizes the null hypothesis that a unit root is present within the time series, indicating non-stationarity [23].

The test equation is formalized as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t \quad (1)$$

where Δy_t represents the difference of the time series at time t , α is the intercept term, βt captures the linear trend, and γ represents the lagged level term. The coefficients δ_1 to δ_{p-1} correspond to the lagged difference terms, with p denoting the number of lags, and ϵ_t is the error term, capturing the deviations not explained by the model.

The ADF test outcome hinges on the test statistic value; a value that is more negative compared to the threshold suggests a stronger likelihood of stationarity within the time series, thereby enhancing the validity of subsequent SARIMAX forecasting [24].

3.3. Temperature Forecasting Using SARIMAX

The Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX) model represents an evolution of the ARIMA model, specifically designed to tackle time series data characterized by seasonal fluctuations and the influence of external factors. This advancement allows the SARIMAX model to encapsulate complex behaviors within time series data, which traditional ARIMA models might overlook, by incorporating exogenous variables into its predictive framework [25]. Central to the SARIMAX model is the SARIMA component, which methodically accounts for the seasonal aspects of the data. The SARIMA notation, denoted as $SARIMA(p,d,q)(P,D,Q)_s$, elucidates the structure of this model component, highlighting its capacity to capture both non-seasonal and seasonal dynamics within the dataset:

- p, d, q represent the non-seasonal components of the model, indicating the order of autoregressive terms, the degree of differencing, and the order of moving average terms, respectively.
- P, D, Q detail the seasonal elements, specifying the order of seasonal autoregressive terms, the degree of seasonal differencing, and the order of seasonal moving average terms.
- s denotes the seasonal periodicity, defining the cycle's length within the time series data.

3.3.1. SARIMAX

The SARIMAX model extends the SARIMA framework by incorporating eXogenous (external) variables, which are additional predictors that can significantly enhance the model's forecasting accuracy. The model can be succinctly expressed as:

$$SARIMAX(p,d,q)(P,D,Q)_s = SARIMA(p,d,q)(P,D,Q)_s + eXogenous\ variables \quad (2)$$

This incorporation of exogenous variables allows the SARIMAX model to accommodate the influences of external factors, providing a more holistic approach to forecasting [26]. The components that constitute the SARIMAX model are as follows:

- **AutoRegressive (AR) Component:** Encapsulates the influence of preceding values on the current value, denoted as $AR(p)$, where p is the number of lag observations included in the model. The AR part is formulated as:

$$X_t = \sum_{n=1}^p \phi_n X_{t-n} + \epsilon_t \quad (3)$$

- **Integrated (I) Component:** Facilitates stationarity in the series by differencing the data, represented as $I(d)$, with d indicating the degree of differencing required.

$$I(d) = X_t - X_{t-d} \quad (4)$$

- **Moving Average (MA) Component:** Models the error of the time series as a linear combination of error terms from previous forecasts, expressed as MA(q), where q is the order of the moving average term [27].

$$X_t = \sum_{n=1}^q \theta_n \epsilon_{t-n} + \epsilon_t \quad (5)$$

- **Seasonal Components:** These include Seasonal AR(P) and Seasonal MA(Q) terms, adding layers to capture seasonal effects within the data.
- **eXogenous Variables (Z_t):** Represents external variables influencing the time series, incorporated as additional predictors to enhance the model's accuracy [28].

The specific settings used for the SARIMAX parameters in our analysis were determined through extensive diagnostic testing to ensure optimal model performance:

- **Parameter settings:** $p = 2$, $d = 1$, $q = 2$ for the non-seasonal components, and $P = 1$, $D = 1$, $Q = 1$, $s = 12$ for the seasonal components, tailored to capture the annual cycle evident in temperature data.
- **Exogenous variables (Z_t):** To enhance the predictive power of the SARIMAX model, we included specific climatic and economic indicators known to affect temperature variations. These exogenous variables were selected based on their relevance and data availability:
 - *Pollution indices:* Including local air quality measurements such as PM2.5 and NO2 concentrations, which have been shown to correlate with temperature anomalies due to their impact on atmospheric composition and heat retention.
 - *Urban development rates:* Quantified through changes in land use patterns, population density, and construction activity within the region. These metrics are sourced from the municipal urban development reports and satellite imagery analysis, reflecting the urban heat island effect which significantly impacts local temperature patterns.
 - *Vegetation indexes:* Utilizing NDVI (Normalized Difference Vegetation Index) data derived from satellite images to account for changes in land cover and their effects on the local climate. Vegetation affects local temperature through evapotranspiration and provides cooling which can be a critical variable in urban settings.
 - *Economic indicators:* Economic growth rates and industrial activity levels, obtained from government economic reports, which indirectly affect temperature through energy consumption patterns and resultant heat emissions.

These variables were integrated into the SARIMAX model to provide a comprehensive view of the factors influencing temperature trends, ensuring that the forecasts account for both direct meteorological conditions and the broader environmental and economic context.

By weaving together these components, the SARIMAX model establishes a robust framework capable of addressing the nuances of seasonal time series data influenced by external variables:

$$X_t = AR + I + MA + SeasonalComponents + Z_t \quad (6)$$

where Z_t represents the exogenous variables, enriching the forecasting model with additional data points that might influence the time series beyond its own historical values.

A deeper examination of the SARIMAX model's architecture reveals its capacity to synthesize various components of time series analysis into a singular predictive framework. This synthesis is articulated through the equation:

$$X_t = \sum_{n=1}^p \phi_n X_{t-n} + \sum_{n=1}^q \theta_n \epsilon_{t-n} + \epsilon_t + (X_t - X_{t-d}) + S(m) + Z_t \quad (7)$$

Each term within this equation plays a pivotal role in modeling the time series data:

- ϕ_i : These autoregressive coefficients quantify the influence of prior values within the series on the current value, embedding the model's memory of past observations.
- θ_i : Moving average coefficients model the impact of past errors (or shocks) on the current observation, allowing the model to adjust for anomalies or unexpected changes in the time series.
- X_t : Represents the actual value of the series at time t , serving as both the target for prediction and a component of the model's calculations.
- ϵ_t : The error term at time t captures the difference between the observed values and those predicted by the model, representing the unexplained variance.
- X_{t-d} : Denotes the value of the series at a time lagged by d periods, essential for the model's differencing process aimed at achieving stationarity.
- $S(m)$: Refers to the seasonal ARMA components, which are crucial for capturing and modeling the cyclical patterns inherent in the data, ensuring that the model accurately reflects seasonal variations.
- Z_t : eXogenous (or external) variables are included as additional predictors to account for the influence of outside factors on the time series. These variables can significantly enhance the model's accuracy by integrating relevant external information, such as economic indicators or environmental factors, that may impact the series.

3.3.2. Seasonal Hyper-Parameter Tuning Using AIC

The Akaike Information Criterion (AIC) is a mathematical technique employed to determine the fit of a model to the data it is derived from. It is instrumental in estimating the optimal values of both ARIMA and seasonal parameters (p, d, q) (P, D, Q), ensuring the model is well-fitted without being overfitted. A lower AIC value is indicative of a model that better fits the data [29]. The AIC is computed based on the number of independent variables K used to construct the model and its capacity to replicate the observed data.

A model achieving a high explanatory power with the fewest independent variables is considered optimal. When comparing models that account for a similar amount of variability, the model with fewer parameters—and consequently, a lower AIC score—is preferred [30,31]. This approach facilitates a comparative evaluation of model quality rather than an absolute measure.

The AIC for a model is calculated using the equation:

$$AIC = -2 \times (\log -likelihood) + 2K \quad (8)$$

where the $\log -likelihood$ component assesses the model's fit to the data, with higher values indicating better fit. The term $2K$, where K is the number of model parameters, penalizes complexity, discouraging the overfitting that can come from models with excessive parameters.

In the context of SARIMAX, the AIC serves a pivotal role in determining the optimal set of parameters—both the ARIMA (p, d, q) and the seasonal (P, D, Q) components. By iterating over possible combinations of these parameters and calculating the AIC for each, the model selection process gravitates towards configurations that offer a judicious blend of predictive accuracy and model parsimony. A lower AIC value signifies a model that has achieved a commendable fit to the data without resorting to unnecessary complexity. This principle of parsimony is crucial, especially in the domain of time series forecasting, where the temptation to "overfit" to historical data can detract from a model's predictive power in unseen scenarios.

The deliberate tuning of SARIMAX parameters guided by the AIC thus represents a methodical approach to enhancing forecasting performance. It is a process of refinement, where the ultimate goal is not just to fit the historical data as closely as possible but to construct a model that generalizes well to future data points. This approach underlines the essence of forecasting as a forward-looking endeavor, where the true measure of a model's value lies in its ability to anticipate future trends with clarity and confidence.

In summary, the AIC's role in SARIMAX modeling transcends mere numerical evaluation. It embodies a principle of modeling efficiency that is crucial for predictive analytics. By steering the

model selection process towards configurations that offer an optimal balance between accuracy and simplicity, AIC-guided hyper-parameter tuning enhances the reliability and robustness of temperature forecasts. This methodological rigor ensures that the SARIMAX model, with its nuanced understanding of seasonal patterns and external influences, stands as a formidable tool in the arsenal of climate prediction and analysis.

3.4. Proposed Algorithm

The endeavor to forecast temperature with high accuracy is critical across various sectors, impacting everything from agricultural productivity to urban climate management. The complex interplay of factors affecting atmospheric conditions necessitates a sophisticated approach to model development. Our proposed methodology leverages the SARIMAX model, an advanced iteration of time series forecasting models, to capture the nuanced effects of seasonality and external variables on temperature trends. Algorithm 1 embodies a structured pathway from data acquisition to comprehensive model evaluation, aimed at predicting the monthly average maximum temperature in Ahmedabad with unprecedented precision.

Algorithm 1 Forecasting Monthly Average Maximum Temperature Using SARIMAX

1: Collect the dataset from the India Meteorological Department (IMD), ensuring a rich and reliable foundation for analysis.

2: Pre-process the data, tailoring it to meet the specific needs of the SARIMAX model. This includes cleaning, normalization, and preparation of time series data.

3: Conduct a stationarity check using the Augmented Dickey-Fuller Test, a critical step to validate the time series' suitability for SARIMAX modeling.

4: **if** p-value < 0.05 **then**

5: Accept the null hypothesis (data is stationary)

6: **else**

7: Perform differencing and return to stationarity check

8: **end if**

9: Apply the statistical time-series based model (SARIMAX)

10: Use the auto-ARIMA model to determine optimal (p, d, q) and (P, D, Q) based on AIC value

11: Forecast the temperature for the required year using data from the previous 10 years

12: Evaluate the performance using MSE, RMSE, R^2 , and MAE

This algorithm represents a strategic synthesis of statistical techniques and model optimization processes designed to fine-tune the forecasting model for maximal accuracy. The forecasting objective, delineated as T_{pred} , encompasses a comprehensive set of monthly temperature projections that are instrumental for both immediate response and long-term strategic planning in various sectors affected by climate variability.

$$T_{pred} = \{T_{m1}, T_{m2}, \dots, T_{m12}\}$$

(9)

where T_{pred} embodies the forecasted monthly average temperatures, each element corresponding to a month's projection, predicated on the analysis of temperature patterns over the preceding decade. This predictive endeavor harnesses the SARIMAX model's capability to dissect and model temperature trends, factoring in both inherent seasonal shifts and the impact of exogenous variables.

The SARIMAX model stands out for its integrative approach, accommodating the complex dynamics of temperature data through its multifaceted components. By conducting rigorous data preprocessing and ensuring data stationarity, the model builds on a solid analytical foundation. The strategic application of auto-ARIMA for parameter selection—guided by the principle of minimizing the AIC—reflects a commitment to model efficiency and accuracy.

Evaluating the model's performance through established metrics allows for a nuanced understanding of its predictive capacity, highlighting areas of strength and potential for further refinement.

This meticulous process not only enhances the reliability of temperature forecasts for Ahmedabad but also sets a precedent for the application of SARIMAX in broader climatological research and operational forecasting.

To sum up, the proposed algorithm encapsulates a comprehensive, methodically structured approach to temperature forecasting. It underscores the SARIMAX model's adaptability and effectiveness in capturing the intricate patterns of temperature variability, offering valuable insights for future research and practical applications in temperature forecasting. This detailed exposition aims to elucidate the algorithm's components and their collective role in advancing the precision and applicability of temperature predictions.

3.4.1. Case 1: Initial Forecasting Approach

Our initial foray into forecasting Ahmedabad's temperature patterns commenced with an auto-ARIMA-driven analysis. Auto-ARIMA, an automated version of the ARIMA model selection process, scans through a predefined range of model parameters—(p, d, q) for the autoregressive, differencing, and moving average components, respectively, and (P, D, Q) for their seasonal counterparts. The goal was to pinpoint the optimal set of parameters that minimizes the Akaike Information Criterion (AIC), a statistical measure used to compare models by balancing goodness of fit with model complexity.

The SARIMAX model, enhanced with the selected parameters, underwent rigorous training. This phase was crucial for assimilating the temporal and seasonal dynamics encapsulated in the dataset, covering a decade of temperature observations. The model's predictive accuracy was initially promising, demonstrating a high degree of alignment with actual temperature data for up to three years post-training. However, the model's performance began to diverge as predictions extended further into the future. This deviation is attributed to the model's reliance on cumulative historical data, where each year's forecasted values were integrated back into the dataset for future predictions. Such a methodology inadvertently magnified the margin of error with each successive forecast. The compounded inaccuracies highlight a fundamental challenge in long-term forecasting: the inherent difficulty of maintaining accuracy over extended periods, especially in the face of fluctuating climate variables and potential shifts in underlying environmental patterns [32].

The escalating forecast uncertainty is illustrated by the broadening error margins observed across different forecast periods. Initially, the model maintained an error range from -4°C to 4°C , which significantly widened to -8°C to 8°C as the forecasting horizon moved from the period of 2003 to 2012 to 2013 to 2023. These findings suggest that while the auto-ARIMA-driven SARIMAX model exhibits robust short- to medium-term forecasting capabilities, its long-term reliability diminishes due to accumulating forecast errors and the dynamic nature of climatic conditions.

The insights derived from Case 1 underscore the necessity for adaptive forecasting strategies capable of accommodating evolving climatic patterns and mitigating the accumulation of predictive errors. Periodic recalibration of the model parameters, informed by continuous data monitoring and validation against emerging climatic trends, may serve as a vital approach to sustaining forecast accuracy over extended periods. Moreover, exploring alternative forecasting models or hybrid approaches that can dynamically adjust to new data inputs and unforeseen environmental variables may offer pathways to enhance the robustness and reliability of long-term temperature forecasts.

In summary, the initial forecasting approach provides a valuable foundation for understanding the operational strengths and limitations of employing auto-ARIMA for SARIMAX model parameter selection. The lessons learned pave the way for developing more resilient forecasting methodologies that can effectively navigate the complexities of long-term climate prediction, ensuring that models remain pertinent and accurate in the face of an ever-changing environmental landscape.

3.4.2. Case 2: Dynamic Forecasting Strategy

Our exploration in Case 2 pioneers a dynamic and iterative forecasting strategy, leveraging the rich temporal dataset spanning from 1993 to 2022 for Ahmedabad. This method initiates with the

model being trained on the initial decade’s data, aiming to predict the immediate subsequent year’s temperatures. Significantly, this procedure is iterative; it involves progressively advancing the training window by one year at each step, thereby ensuring that the model is continuously refreshed with the most recent data, enhancing its predictive accuracy over time.

One of the cornerstone decisions in this methodology is the intentional exclusion of forecasted values from the subsequent training sets. This strategic choice is predicated on the understanding that forecasted values, while informative, inherently possess a degree of uncertainty. By relying exclusively on observed data for model training, we ensure that the integrity and accuracy of the forecasting process are preserved, anchoring the model’s predictions in empirical evidence. The adoption of a rolling-window training approach underpins the model’s robustness and adaptability. This methodology ensures that the model is not only perpetually updated with the latest data but also rigorously tested against diverse data subsets. This reflective approach closely mirrors real-world forecasting conditions, thereby enhancing the model’s ability to adapt to evolving climatic patterns and improving its generalization capability over an extended temporal horizon.

The efficacy of this dynamic forecasting strategy is vividly illustrated through the comparison of the model’s predictions against actual temperature values for two distinct periods, 2003-2013 and 2014-2023, as depicted in Figures 2(a) and 2(b). These visual representations offer a compelling testament to the model’s predictive precision and its capacity to adjust to temporal shifts in climatic trends over time, affirming its utility and effectiveness in the domain of climate forecasting.

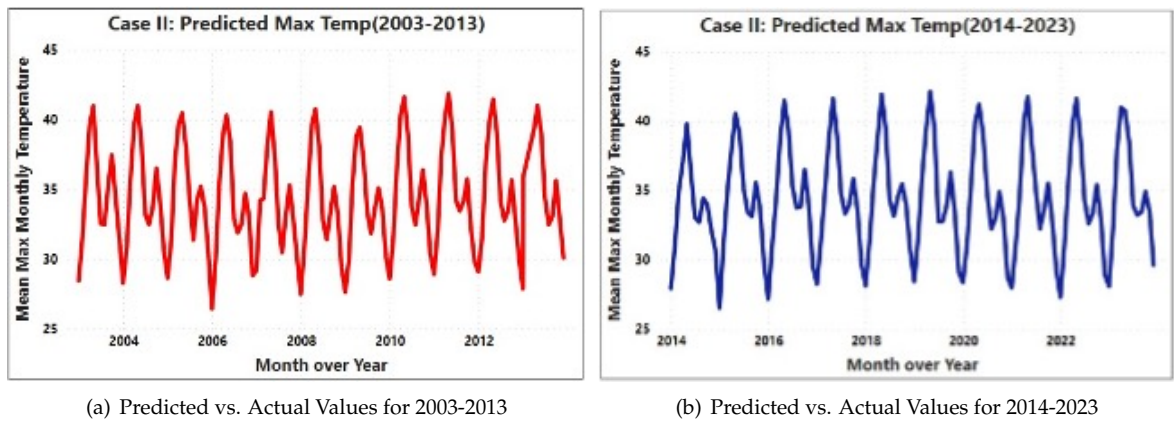


Figure 2. Comparison of the model’s predicted values against actual temperatures, illustrating the dynamic forecasting strategy’s efficacy in adapting to temporal shifts in climate patterns.

The unfolding of Case 2 showcases not just a methodological advancement in temperature prediction but also a paradigmatic shift towards a more adaptive and responsive forecasting framework. This strategy exemplifies the imperative for climate models to be dynamic, capable of evolving in tandem with their datasets, to maintain relevance and accuracy in an ever-changing environmental landscape.

Looking ahead, the implications of this approach are profound, offering a blueprint for further enhancing the forecasting model. The potential integration of cutting-edge machine learning algorithms or the inclusion of a broader array of climatic and environmental variables could substantially refine predictive capabilities. Furthermore, the successful application of this method to different geographical regions or varying climatic phenomena opens new avenues for research and innovation in global climate prediction, underscoring the pivotal role of adaptability and iterative learning in advancing our understanding and management of climate variability.

3.4.3. Comparative Analysis of Case 1 and Case 2

In the comparative analysis between Case 1 and Case 2, we delve into the nuances that delineate their distinct forecasting strategies and outcomes. Case 1 leverages an auto-ARIMA function to deduce optimal SARIMAX model parameters, with a focused strategy on minimizing the Akaike Information Criterion (AIC) for enhanced model accuracy in the short to medium term. Initially, this methodology demonstrates promising forecast precision, aligning closely with actual data for the initial years following model training. However, the predictive accuracy begins to diverge as the forecasting horizon extends, revealing a fundamental limitation associated with the model's dependency on historical data. This reliance initiates an error accumulation process, which, compounded by unforeseen climatic events and shifts in environmental patterns, results in a significant expansion of the error range over time.

Contrasting sharply with Case 1, Case 2 adopts a dynamic and iterative forecasting strategy through a rolling-window training approach. This method, covering nearly three decades of data, continuously adapts to evolving patterns, significantly enhancing the model's predictive accuracy and generalization capability across a prolonged timeframe. A pivotal decision in this strategy is to eschew the inclusion of predicted values in the training dataset for subsequent iterations, a move designed to prevent the snowballing of forecast errors. This approach, augmented by rigorous evaluation metrics, not only showcases a robust capacity for model adaptation but also maintains the forecasts' empirical grounding, thus ensuring the model's long-term predictive reliability.

The graphical representations of forecasted versus actual temperatures for distinct periods further elucidate the comparative strengths of the models employed in both cases. The fidelity of Case 2's forecasts to actual climatic outcomes, across extended timeframes, accentuates its methodological superiority in capturing and adapting to temporal climatic trends. This alignment, or its absence, constitutes a critical evaluation metric for the model's efficacy, spotlighting areas for potential refinement to further optimize forecasting accuracy.

The juxtaposition of Case 1 and Case 2 elucidates a broader narrative within climate forecasting: the imperative for methodologies that not only yield immediate accuracy but also sustain adaptability over time. Case 2's rolling-window approach embodies this principle, offering a scalable and flexible framework capable of navigating the complexities of climate prediction with enhanced precision. This comparative analysis not only advances our understanding of forecasting models' operational dynamics but also underscores the strategic importance of methodological adaptability in the face of climatic unpredictability.

4. Experimental Evaluation

4.1. Results

In the quest to assess the precision of our monthly average maximum temperature predictions, we employed a suite of metrics to evaluate model performance across three decades [33,34]. The accuracy of the forecasts was quantified using key indicators such as Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and the R-squared (R^2) score, as illustrated in the following equations:

$$\text{RMSE}(T) = \sqrt{\frac{1}{m} \sum_{i=1}^m (g_i - \hat{g}_i)^2} \quad (10)$$

$$\text{MSE}(T) = \frac{1}{m} \sum_{i=1}^m (g_i - \hat{g}_i)^2 \quad (11)$$

$$\text{MAE}(T) = \frac{1}{m} \sum_{i=1}^m |g_i - \hat{g}_i| \quad (12)$$

$$R^2(T) = \frac{\sum_{i=1}^m (g_i - \hat{g}_i)^2}{\sum_{i=1}^m (g_i - \bar{g})^2} \tag{13}$$

where T represents the year, i the month, g_i the actual temperature, \hat{g}_i the predicted temperature, and \bar{g} the average of actual temperatures. This methodology ensures a rigorous assessment of the forecasting model’s effectiveness over an extended period.

The RMSE values for forecasted years from 2003 to 2013 are presented in Table 3, showcasing the model’s performance and offering insights into its temporal reliability.

Table 3. Error Values for Predicted Years (2003 - 2013).

Year Used for Training	Predicted Year	RMSE
1992 - 2002	2003	1.4185
1993 - 2003	2004	1.5347
1994 - 2004	2005	1.4740
1995 - 2005	2006	1.5436
1996 - 2006	2007	1.3115
1997 - 2007	2008	1.0265
1998 - 2008	2009	1.7331
1999 - 2009	2010	1.4287
2000 - 2010	2011	1.1369
2001 - 2011	2012	1.1474
2002 - 2012	2013	1.3161

An in-depth analysis of these values indicates a significant correlation between predicted and actual temperatures, suggesting high predictive accuracy. This is further affirmed by a refined equation:

$$t_{\text{actual}} = t_{\text{pred}} + (M \pm \delta t) \tag{14}$$

where M is a constant value determined to be 1, indicating a systematic underestimation by the model, and δt represents a minor deviation in temperature, ranging between 0.1 and 0.9. This equation acknowledges a consistent bias in the model’s forecasts, suggesting that, despite the low error magnitude (with RMSE values hovering around 1), there exists a predictable deviation from actual temperature values.

The identification of this systematic bias is crucial for further model refinement. Understanding the nature of this bias and exploring avenues for adjustment—be it through calibration of the model’s coefficients or the incorporation of additional predictive variables—can significantly enhance the model’s accuracy. Such enhancements aim to better capture the complex dynamics underlying temperature variations, thereby improving the reliability and applicability of the model’s forecasts in real-world scenarios [35].

4.2. Comparative Performance Analysis of Forecasting Models

This subsection provides a comparative analysis of forecasting models from recent literature, juxtaposing them against the advancements introduced by the SARIMAX model with hyper-parameter tuning in this study. The comparative lens primarily focuses on RMSE, a universally recognized metric for evaluating the accuracy of predictions across various time series forecasting challenges, including temperature prediction.

The WD-SARIMAX method delivered an RMSE score of 1.67 for temperature prediction tasks [11]. This stands in contrast to the refined SARIMAX model proposed in our study, which integrates meticulous hyper-parameter tuning aimed at optimizing the model’s performance. The effectiveness of this enhancement is evidenced by an improved RMSE value of 1.026, underscoring the potential of hyper-parameter optimization in elevating forecasting accuracy.

Temperature forecasting utilizing the Bi-directional LSTM model yielded an RMSE value of 1.74 [14]. This approach did not account for the stationarity of the data, a critical factor in time series analysis. Conversely, our methodology incorporates the Augmented Dicky Fuller Test for stationarity verification, combined with targeted fine-tuning of the model parameters. This strategic adjustment not only improves computational efficiency but also notably refines the RMSE to 1.026, affirming the importance of data pre-processing in enhancing model precision.

In [20], SARIMAX was augmented with external regressors for daily demand forecasting in a hotel context, trained on a dataset covering five years. An interesting revelation from this approach was the identification of non-seasonality through the Augmented Dickey-Fuller (ADF) test, challenging the typical expectation of seasonal patterns in such data. The resultant RMSE of 10.365 starkly contrasts with the outcomes from temperature forecasting, highlighting the divergent nature of data characteristics and forecasting objectives across different domains.

The application of the SARIMA model alongside the Markov Chain Monte Carlo (MCMC) method for wind energy forecasting showcased an average RMSE of 16.44%, predicated on a relatively small dataset of 550 values [36]. The limitation imposed by the dataset size underscores the necessity for ample data in capturing seasonal trends effectively. While MCMC offers enhanced precision for long-range quantitative forecasts, its applicability to temperature forecasting may be constrained by the inherent differences in the nature of the data and prediction objectives.

As we delve into the comparative analysis of various forecasting methodologies, it is pivotal to contextualize the performance of each approach within the framework of predictive accuracy. Table 4 succinctly summarizes the RMSE values obtained from different methods, including the WD-SARIMAX, Bi-directional LSTM, and SARIMAX with external regressors, among others. This table not only highlights the diversity of techniques applied to forecasting challenges but also underscores the proposed SARIMAX methodology’s standout performance. By scrutinizing these RMSE scores, we can glean insights into the efficacy of hyper-parameter tuning and data preprocessing, which are instrumental in the proposed method’s enhanced accuracy.

Table 4. Comprehensive Comparative Analysis of Forecasting Models.

Paper	Methodology	RMSE
[11]	WD-SARIMAX for Temperature Prediction	1.67
[14]	Bi-directional LSTM for Temperature Forecasting	1.74
[19]	Wind Energy Analysis with SARIMA and MCMC	14.66
[20]	Demand Forecasting with SARIMAX	10.365
[21]	Temperature Forecasting with Mann-Kendall and ARIMA	1.40
[22]	Weather Forecasting using Artificial Neural Network	2.55
	Proposed SARIMAX Methodology	1.026

The comparative analysis elucidates the nuanced performance landscape of various forecasting models, emphasizing the significant strides made by the proposed SARIMAX methodology in achieving superior RMSE values. This benchmarking not only highlights the efficacy of incorporating hyper-parameter tuning and data stationarity checks but also positions the proposed model as a robust tool for precise temperature forecasting, setting a new standard for future research and application in the field. In conclusion, by demystifying complex climate data, the proposed model can inform public awareness initiatives and foster participatory approaches in urban environmental management, ensuring that sustainable development is a shared vision.

4.3. Discussion

The experimental evaluation and subsequent comparative analysis underscore the proposed SARIMAX methodology’s enhanced precision in forecasting monthly average maximum temperatures. The meticulous integration of hyper-parameter tuning, alongside the incorporation of stationarity

checks via the Augmented Dicky Fuller Test, marks a significant methodological advancement, as evidenced by the superior RMSE score of 1.026 when compared against other forecasting models.

The observed systematic underestimation by the proposed model, indicated by a consistent deviation in the predictive equation, illuminates an area for further refinement. This bias, while minor, opens discussions on the potential for integrating additional climatic variables or adjusting model parameters to better capture extreme weather events or the gradual effects of climate change on temperature patterns. Furthermore, the comparative analysis reveals the broad spectrum of RMSE scores across different models and applications, from temperature forecasting to energy demand prediction. This variability not only highlights the diverse challenges inherent in time series forecasting but also the critical role of dataset characteristics—such as size and seasonality—in influencing model performance.

The proposed methodology's standout performance, particularly in the context of temperature forecasting, suggests its applicability to a wider range of time series prediction tasks. However, the insights garnered from the comparison with other models, especially those employing different data preprocessing techniques or analytical frameworks, suggest that no single model universally excels across all scenarios. Instead, the choice of model and its configuration should be carefully tailored to the specific characteristics of the dataset and the forecasting objectives at hand.

5. Conclusions and Future Work

This research has contributed to the field of temperature forecasting by developing a sophisticated SARIMAX model, specifically tailored to address the unique climatic nuances of Ahmedabad. The model's adaptability and precision in forecasting monthly average maximum temperatures stand as a testament to its potential utility in environmental planning and policy formulation. By providing accurate temperature predictions, the model equips decision-makers and policymakers with the necessary insights to devise and implement strategies aimed at mitigating the adverse effects of climate change, thus fostering environmental sustainability and enhancing community well-being. Moreover, the study underscores the importance of integrating advanced statistical techniques and hyper-parameter optimization to improve forecasting accuracy. The deliberate incorporation of stationarity checks through the Augmented Dicky Fuller Test further highlights the model's methodological robustness, ensuring reliability in its predictive capabilities.

As we look towards the horizon of temperature forecasting, several avenues emerge for refining the SARIMAX model and broadening the scope of the current research work. A deeper analysis of seasonal patterns and extreme weather events is paramount. By dissecting these temporal fluctuations with greater precision, we can uncover intricate climatic behaviors that the current model may overlook. Enhancing the model's sensitivity to such nuances promises not only to sharpen its forecasting accuracy but also to provide a richer understanding of climate dynamics over time. Furthermore, the inclusion of additional environmental variables—ranging from humidity and precipitation to pollution levels—into the forecasting model could yield a more comprehensive framework for understanding temperature variations. This holistic approach would allow for a more nuanced analysis of the multifaceted influences on climate, facilitating more accurate and meaningful predictions.

Exploring alternative forecasting methodologies presents another fruitful direction. The integration of machine learning algorithms or ensemble models could introduce new dimensions to temperature prediction, potentially offering improvements in both accuracy and efficiency. Such investigative efforts would contribute to a diverse toolkit for climatologists and data scientists, enabling them to select the most appropriate techniques based on the specific challenges and data characteristics at hand. Expanding the geographical scope of the SARIMAX model to encompass various cities, regions, or climatic zones is another critical step forward. This expansion would not only validate the model's adaptability but also pave the way for the development of localized models that cater to the distinct climatic conditions across different parts of the world. Such region-specific models could

significantly enhance the accuracy of local weather forecasts, providing valuable insights for regional planning and disaster management.

Incorporating mechanisms for dynamic model updating based on real-time data streams could dramatically improve the model's adaptability to changing weather conditions. This approach would ensure that the forecasting model remains relevant and reliable, even as rapid environmental changes occur [32]. By continuously updating its parameters in response to the latest data, the model could offer more accurate predictions, adjusting swiftly to unforeseen climatic shifts. Embarking on all these paths of inquiry and development, we aim to fortify the foundations of temperature forecasting. The ultimate goal is to harness advanced predictive analytics to better prepare for and mitigate the impacts of climate variability, ensuring a sustainable and resilient future for communities worldwide.

From a broad perspective, the current research contributes to the global agenda of achieving SDG11 by highlighting how sophisticated temperature forecasting can support the sustainable management of urban landscapes. Looking ahead, integrating the proposed approach with cutting-edge technologies such as artificial intelligence, geospatial analysis, and the metaverse opens new avenues for creating resilient, adaptive, and smart urban environments. Future research will explore these integrations, aiming to provide comprehensive solutions for the multifaceted challenges of sustainable urban development.

References

1. National Oceanic and Atmospheric Administration: Excessive heat, a 'silent killer'. <https://www.noaa.gov/stories/excessive-heat-silent-killer>. Online; accessed on 16 July 2024.
2. Gustin, M.; McLeod, R.S.; Lomas, K.J. Forecasting Indoor Temperatures during Heatwaves using Time Series Models. *Building and Environment* **2018**, *143*, 727–739.
3. The Intergovernmental Panel on Climate Change (IPCC). https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM_final.pdf. Online; accessed on 16 July 2024.
4. Data Dive: Land Lost to Forest Fires in India Increases by 122 <https://www.factchecker.in/data-dive/data-dive-land-lost-to-forest-fires-in-india-increases-by-122-in-5-years-815025>. Online; accessed on 16 July 2024.
5. The Climate Action Button. <https://climatebutton.ucsusa.org/>. Online; accessed on 16 July 2024.
6. Kreuzer, D.; Munz, M.; Schlüter, S. Short-term temperature forecasts using a convolutional neural network—An application to different weather stations in Germany. *Machine Learning with Applications* **2020**, *2*, 100007.
7. Sustainable Development Goals: 17 Goals to Transform our World - Ensure access to affordable, reliable, sustainable and modern energy. <https://www.un.org/sustainabledevelopment/energy/>. Online; accessed on 16 July 2024.
8. Veeramsetty, V.; Kiran, P.; Sushma, M.; Salkuti, S.R. Weather Forecasting Using Radial Basis Function Neural Network in Warangal, India. *Urban Science* **2023**, *7*, 68.
9. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. 35th AAAI Conference on Artificial Intelligence (AAAI). AAAI Press, 2021, pp. 11106–11115.
10. Roy, D.S. Forecasting The Air Temperature at a Weather Station Using Deep Neural Networks. *Procedia Computer Science* **2020**, *178*, 38–46.
11. Elshewey, A.M.; Shams, M.Y.; Elhady, A.M.; Shohieb, S.M.; Abdelhamid, A.A.; Ibrahim, A.; Tarek, Z. A Novel WD-SARIMAX Model for Temperature Forecasting Using Daily Delhi Climate Dataset. *Sustainability* **2022**, *15*, 757.
12. Cadenas, E.; Rivera, W.; Campos-Amezcu, R.; Heard, C. Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model. *Energies* **2016**, *9*, 109.
13. Hewage, P.; Trovati, M.; Pereira, E.; Behera, A. Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications* **2021**, *24*, 343–366.
14. Zenkner, G.; Navarro-Martinez, S. A Flexible and Lightweight Deep Learning Weather Forecasting Model. *Applied Intelligence* **2023**, *53*, 24991–25002.

15. Thakur, N.; Karmakar, S.; Soni, S. Time Series Forecasting for Uni-variant Data using Hybrid GA-OLSTM Model and Performance Evaluations. *International Journal of Information Technology* **2022**, *14*, 1961–1966.
16. Biswas, M.; Dhoom, T.; Barua, S. Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data. *International Journal of Computer Applications* **2018**, *182*, 20–24.
17. U, J.K.; Kovoov, B.C. Deterministic Weather Forecasting Models based on Intelligent Predictors: A Survey. *Journal of King Saud University - Computer and Information Sciences* **2022**, *34*, 3393–3412.
18. Aslam, M. Time Series Data Analysis under Indeterminacy. *Journal of Big Data* **2023**, *10*, 126.
19. Al-Duais, F.S.; Al-Sharpi, R.S. A Unique Markov Chain Monte Carlo Method for Forecasting Wind Power Utilizing Time Series Model. *Alexandria Engineering Journal* **2023**, *74*, 51–63.
20. Ampountolas, A. Modeling and Forecasting Daily Hotel Demand: A Comparison Based on SARIMAX, Neural Networks, and GARCH Models. *Forecasting* **2021**, *3*, 580–595.
21. Kaur, B.; Kaur, N.; Gill, K.K.; Singh, J.; Bhan, S.C.; Saha, S. Forecasting Mean Monthly Maximum and Minimum Air Temperature of Jalandhar District of Punjab, India using Seasonal ARIMA Model. *Journal of Agrometeorology* **2022**, *24*, 42–49.
22. Abhishek, K.; Singh, M.; Ghosh, S.; Anand, A. Weather Forecasting Model using Artificial Neural Network. *Procedia Technology* **2012**, *4*, 311–318.
23. Ajewole, K.P.; Adejuwon, S.O.; Jemilohun, V.G. Test for Stationarity on Inflation Rates in Nigeria using Augmented Dickey Fuller Test and Phillips-Persons Test. *IOSR Journal of Mathematics* **2020**, *16*, 11–14.
24. Zhang, Z.; Dong, Y. Temperature Forecasting via Convolutional Recurrent Neural Networks Based on Time-Series Data. *Complexity* **2020**, *2020*, 3536572:1–3536572:8.
25. Brown, G.D.; Largey, A.; McMullan, C.; Reilly, N.; Sahdev, M. Weathering the Storm: Developing a User-centric Weather Forecast and Warning System for Ireland. *International Journal of Disaster Risk Reduction* **2023**, *91*, 103687.
26. Yang, B.; Ma, T.; Huang, X. ATFSAD: Enhancing Long Sequence Time-Series Forecasting on Air Temperature Prediction. *IEEE Access* **2023**, *11*, 92080–92091.
27. Lynch, P. The Origins of Computer Weather Prediction and Climate Modeling. *Journal of Computational Physics* **2008**, *227*, 3431–3444.
28. Scher, S.; Messori, G. Predicting Weather Forecast Uncertainty with Machine Learning. *Quarterly Journal of the Royal Meteorological Society* **2018**, *144*, 2830–2841.
29. Chai, T.; Draxler, R.R. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature. *Geoscientific Model Development* **2014**, *7*, 1247–1250.
30. Merabet, K.; Heddarn, S. Improving the Accuracy of Air Relative Humidity Prediction using Hybrid Machine Learning based on Empirical Mode Decomposition: A Comparative Study. *Environmental Science and Pollution Research* **2023**, *30*, 60868–60889.
31. Tao, H.; Awadh, S.M.; Salih, S.Q.; Shafik, S.S.; Yaseen, Z.M. Integration of Extreme Gradient Boosting Feature Selection Approach with Machine Learning Models: Application of Weather Relative Humidity Prediction. *Neural Computing and Applications* **2022**, *34*, 515–533.
32. Swain, D.; Vijeta.; Manjare, S.; Kulawade, S.; Sharma, T. Stock Market Prediction Using Long Short-Term Memory Model. *Machine Learning and Information Processing*, 2021, pp. 83–90.
33. Amnuaylojaroen, T. Advancements in Downscaling Global Climate Model Temperature Data in Southeast Asia: A Machine Learning Approach. *Forecasting* **2023**, *6*, 1–17.
34. Shrivastav, L.K.; Jha, S.K. A Gradient Boosting Machine Learning Approach in Modeling the Impact of Temperature and Humidity on the Transmission Rate of COVID-19 in India. *Applied Intelligence* **2021**, *51*, 2727–2739.

35. Zohdi, M.; Rafiee, M.; Kayvanfar, V.; Salamiraad, A. Demand Forecasting based Machine Learning Algorithms on Customer Information: An Applied Approach. *International Journal of Information Technology* **2022**, *14*, 1937–1947.
36. Bojer, C.S. Understanding Machine Learning-based Forecasting Methods: A Decomposition Framework and Research Opportunities. *International Journal of Forecasting* **2022**, *38*, 1555–1561.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.