



Weather Forecasting and Climate Analytics Using R

Oshini Jadhav¹, Sanika Sontake², Ishwari Waghule³, Prof. Roshni Kolhe⁴, Dr. Bharti Sayankar⁵, Dr. Sudhir N. Shelke⁶

¹ Department of CSE , Guru Nanak Institute of Technology ,Nagpur, India

² Department of CSE , Guru Nanak Institute of Technology ,Nagpur, India

³ Department of CSE , Guru Nanak Institute of Technology ,Nagpur, India

⁴ Department of CSE , GNIT (Project Guide), Nagpur, India

⁵ Department of CSE , GNI(Head of Department),Nagpur, India

⁶ Principal & Director of (GNIT), Nagpur, India

Corresponding Author Email: oshinijadhav@gmail.com

How to Cite this Article:

Jadhav, O., Sontake, S., Waghule, I., Kolhe, R. & Sayankar, B. (2026). Weather Forecasting and Climate Analytics Using R. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(10), 1-9.
<https://doi.org/10.55041/ijcope.v2i10.060>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.060>

Abstract—

Weather conditions vary continuously and are influenced by multiple atmospheric parameters, making the analysis and prediction of weather patterns an important area of data-driven research. This paper presents a study of a Weather Forecasting and Climate Analytics system developed using R programming. The system focuses on the analysis of historical weather data containing parameters such as temperature, humidity, rainfall, wind speed, and atmospheric pressure. The methodology includes data preprocessing, exploratory data analysis, visualization, correlation analysis, and time-series analysis to identify important patterns and relationships within the dataset. The Autoregressive Integrated Moving Average (ARIMA) model is applied to forecast future weather conditions based on historical observations and temporal patterns. In addition to forecasting, the proposed system focuses on climate analytics by examining monthly and seasonal variations, rainfall patterns, temperature trends, humidity behavior, and weather differences across different locations. Graphical representations are used to make these patterns easier to interpret and understand. The proposed approach combines statistical analysis, visualization, and ARIMA-based forecasting within the R environment to provide an integrated framework for understanding historical weather behavior and generating data-driven predictions. The study also considers limitations related to dataset quality, missing observations, extreme weather events, and the ability of traditional time-series models to capture complex atmospheric relationships. The approach can provide useful analytical support for applications such as environmental monitoring,

agriculture, and weather-related planning.

Keywords— Weather Forecasting; Climate Analytics; R Programming; Time-Series Analysis; ARIMA.

I. INTRODUCTION

Weather forecasting plays an important role in supporting decision-making in agriculture, transportation, environmental monitoring, water-resource management, disaster preparedness, and other weather-sensitive activities. Weather conditions are dynamic and are influenced by several interconnected

atmospheric variables, including temperature, rainfall, humidity, wind speed, and atmospheric pressure. The increasing availability of historical meteorological data has created opportunities to analyze these variables using statistical and computational techniques and to identify temporal patterns that can support future weather prediction.



Conventional weather forecasting relies extensively on meteorological observations and numerical weather prediction approaches based on physical and mathematical representations of atmospheric processes. In parallel with these methods, statistical time-series techniques provide a data-driven approach for analyzing historical observations and forecasting future values. The Autoregressive Integrated Moving Average (ARIMA) model is a widely used statistical approach for time-dependent data and combines autoregressive, differencing, and moving-average components to model temporal relationships. ARIMA is particularly useful when historical observations contain identifiable temporal dependencies and trends.

The analysis of historical weather data is not limited to predicting future values. Understanding long-term and seasonal variations in temperature, rainfall, humidity, and other meteorological parameters can provide useful information about weather and climate behavior. Data preprocessing, exploratory data analysis, visualization, and correlation analysis can help identify patterns and relationships among variables before forecasting models are applied. Therefore, an integrated analytical approach can provide a broader understanding of weather data rather than focusing only on future predictions.

However, weather datasets may contain missing observations, inconsistent records, temporal variations, seasonal behavior, and relationships among multiple meteorological parameters. Furthermore, forecasting performance can vary depending on the characteristics of the dataset, the forecasting horizon, and the selected modeling technique. Existing research has investigated statistical, machine-learning, and deep-learning approaches for weather and environmental forecasting, but different methods can perform differently under different data conditions. This creates a need for systematic analysis of historical weather data, appropriate preprocessing, visualization, and selection of suitable forecasting techniques.

The present study addresses this requirement through a Weather Forecasting and Climate Analytics system using R. The proposed approach integrates historical weather-data preprocessing, exploratory data analysis, visualization, correlation analysis, climate-pattern analysis, and ARIMA-based time-series forecasting within a single analytical framework. The project focuses on weather parameters such as temperature, humidity, rainfall, wind speed, and atmospheric pressure. This integrated approach is intended to

provide both an understanding of historical weather patterns and a data-driven basis for forecasting future conditions.

The main objectives of the study are to: (i) preprocess and organize historical weather data; (ii) analyze important meteorological parameters through exploratory data analysis; (iii) visualize weather trends and variations; (iv) examine relationships among weather variables using correlation analysis; (v) identify trends and seasonal patterns through time-series analysis; (vi) apply the ARIMA model for weather forecasting; and (vii) evaluate forecasting performance using appropriate error metrics. Through these objectives, the study aims to develop a structured R-based framework for combining weather analytics and ARIMA-based forecasting.

II. LITERATURE REVIEW

Weather forecasting and environmental time-series analysis have been studied using a wide range of statistical, machine-learning, and deep-learning techniques. Existing research demonstrates that the suitability of a forecasting model depends on the characteristics of the weather variable, the temporal resolution of the dataset, the forecasting horizon, the presence of seasonal patterns, and the quantity and quality of available observations. The following studies provide a foundation for understanding the development of statistical and data-driven approaches to weather forecasting.

A. Statistical Time-Series Approaches

Statistical time-series models have traditionally been used to forecast environmental variables because they explicitly model temporal relationships within historical observations. ARIMA is one of the commonly applied statistical approaches and combines autoregressive terms, differencing, and moving-average components to model time-dependent data. Lai and Dzombak [1] investigated the use of ARIMA for near-term regional temperature and precipitation projections and reported that the approach could provide interpretable and useful location-specific forecasts. However, their results also showed that ARIMA did not outperform every alternative method for all climate indices, indicating that model performance depends on the characteristics of the variable and dataset [1].

The application of ARIMA to Indian meteorological data has also been investigated. Gill et al. studied temperature and rainfall time-series prediction in



Punjab using ARIMA-based approaches and considered seasonal forecasting methods for meteorological data exhibiting recurring patterns [2]. Such studies demonstrate the relevance of statistical time-series models for regional weather analysis and provide a foundation for applying similar approaches to historical weather datasets.

B. Seasonal Forecasting Using SARIMA

Weather observations frequently contain seasonal patterns, particularly in variables such as rainfall and temperature. SARIMA extends the ARIMA framework by incorporating seasonal autoregressive, differencing, and moving-average components. This makes it suitable for time series in which observations exhibit recurring behaviour at regular intervals.

A 2024 study on SARIMA-based weather forecasting investigated the use of the model for weather prediction and emphasized its ability to represent both trend and seasonal components in historical weather data [3]. The study demonstrates the potential usefulness of seasonal statistical models when weather observations contain periodic behaviour. However, the effectiveness of SARIMA remains dependent on the quality and structure of the input data and the selection of appropriate model parameters.

C. Machine Learning and Deep Learning Approaches

The development of machine-learning and deep-learning techniques has expanded the range of approaches available for environmental forecasting. Unlike conventional statistical models, neural-network-based methods can model more complex nonlinear relationships when sufficient training data are available. Long Short-Term Memory (LSTM) networks, in particular, have been widely investigated for sequential and time-dependent prediction because they are designed to retain information across time steps.

Comparative research has examined the differences between ARIMA and LSTM for environmental variables. A study on wind-speed forecasting compared ARIMA with artificial neural-network and LSTM approaches and reported lower forecasting error for LSTM in its experimental setting. At the same time, the study noted that the performance of traditional statistical approaches can depend on the quantity and characteristics of the available data [4].

Similarly, research on rainfall forecasting has compared ARIMA and LSTM approaches using environmental datasets, reflecting the need to evaluate both traditional statistical and modern deep-learning techniques for precipitation prediction [5]. These studies indicate that there is no universally applicable forecasting technique; instead, model suitability depends on factors such as data characteristics, temporal structure, forecasting horizon, and computational requirements.

D. Comparison of Existing Approaches

The literature indicates several important differences between statistical and learning-based forecasting approaches. ARIMA and SARIMA provide relatively interpretable statistical frameworks and can be effective when temporal dependencies and seasonal patterns are identifiable. They can also be implemented with comparatively lower computational requirements than deep-learning approaches. On the other hand, LSTM and related neural-network models are capable of representing nonlinear temporal relationships but generally require more data, model training, and parameter tuning.

Therefore, model complexity should not be considered the only criterion for selecting a forecasting technique. Research comparing ARIMA and LSTM has shown that the relative performance of the models can change according to the dataset and forecasting task. For example, studies in environmental and related time-series domains have reported different outcomes when comparing classical statistical models with LSTM-based approaches [4], [5].

E. Research Gap

The reviewed studies demonstrate substantial work on individual forecasting techniques, particularly ARIMA, SARIMA, and machine-learning or deep-learning models. However, much of the existing research is focused primarily on forecasting a particular weather variable or comparing specific forecasting models. Less emphasis is placed on presenting an integrated analytical workflow that combines data preprocessing, exploratory data analysis, visualization, correlation analysis, climate-pattern analysis, and statistical forecasting within a single R-based framework.

The present study addresses this gap by combining weather-data analytics and forecasting in an integrated system. Instead of considering forecasting as an isolated step, the proposed approach first examines the quality and characteristics of historical weather data, identifies



relationships and temporal patterns, and then applies appropriate time-series forecasting techniques. This provides a broader analytical framework in which historical climate behaviour and future weather prediction can be studied together.

Overall, the literature supports the use of time-series methods for weather and environmental forecasting while also demonstrating the growing importance of machine-learning and deep-learning alternatives. The reviewed research therefore provides the theoretical and methodological basis for the proposed R-based Weather Forecasting and Climate Analytics system.

III. METHODOLOGY

A. Research Design

This study uses a data-driven approach for weather forecasting and climate analytics using R. The methodology includes data preprocessing, exploratory data analysis, correlation analysis, time-series analysis, visualization, and ARIMA-based forecasting. The complete analysis was performed using RStudio.

B. Data Collection

The dataset, **weather forecasting dataset**, was obtained from Kaggle. It contains **3,655 records and 7 columns** covering five cities: **Nagpur, Mumbai, Pune, Hyderabad, and Delhi**. The variables are Date, City, Temperature C, Humidity pct, Rainfall mm, Wind Speed km/h, and Pressure Hectopascal.

C. Data Preprocessing

The dataset was inspected for missing values, duplicate records, and inconsistencies. The Date column was converted into an appropriate date format and the observations were arranged chronologically. The data was organized by city and time to support further analysis and forecasting.

D. Exploratory and Statistical Analysis

Exploratory Data Analysis was performed to examine temperature, humidity, rainfall, wind speed, and atmospheric pressure. Statistical summaries and graphs were used to identify trends, variations, and seasonal patterns. Correlation analysis was also performed to examine relationships between different weather parameters

E. Time-Series Forecasting

The historical weather observations were converted into time-series data for forecasting. The **Autoregressive Integrated Moving Average (ARIMA)** model was applied to identify temporal patterns and generate future weather predictions. The forecasting results were calculated from historical observations and represented graphically.

F. Climate Analytics and Visualization

Climate analytics was performed to study monthly, seasonal, and city-wise variations in weather conditions. R was used to generate graphs showing trends, relationships, and forecasted values. These

Author/Y ear	Weather Variable	Dataset/L ocation	Metho d	Limitati on
Y. Lai and D. A. Dzombak	Temperat ure, precipitati on	Regional	ARIM A	Not best for every index
Gill et al.	Temperat ure, rainfall	Punjab	ARIM A/SAR IMA	Dataset/ location specific
2024 study	Weather variables	—	SARI MA	Depend ent on input/m odel paramet ers
Wind-speed study	Wind speed	—	ARIM A, ANN, LSTM	Model/d ata depende nt
Rainfall study	Rainfall	Multiple regions	ARIM A, LSTM	Dataset depende nt



visualizations were interpreted to understand historical weather behaviour and differences among the selected cities.

G. Software and Ethical Considerations

The complete implementation was carried out in **RStudio using R programming**. The study uses publicly available weather data from Kaggle and does not involve personal or sensitive information. Therefore, no human-subject or personal-data concerns were involved.

IV. RESULTS AND DISCUSSION

The analysis of the **weather forecasting dataset** provided insights into historical weather patterns across Nagpur, Mumbai, Pune, Hyderabad, and Delhi. The dataset was analysed using RStudio through exploratory analysis, correlation analysis, visualization, and ARIMA-based time-series forecasting.

A. Exploratory Analysis

The exploratory analysis showed variations in temperature, humidity, rainfall, wind speed, and atmospheric pressure over time. Graphical analysis helped identify temporal patterns and differences in weather conditions among the selected cities. These results demonstrate the usefulness of historical weather data for understanding changes in weather parameters.

Table I: Weather Dataset Characteristics

Parameter	Description
Data Source	Kaggle
Dataset	Weather forecasting dataset
Variables	7
Cities	Nagpur, Mumbai, Pune, Hyderabad, Delhi

B. Correlation Analysis

Correlation analysis was used to examine relationships among temperature, humidity, rainfall, wind speed, and atmospheric pressure. The correlation results provided an understanding of how changes in one weather parameter were associated with changes in other parameters. These relationships supported the interpretation of the patterns observed in the graphical analysis.

C. Time-Series Forecasting

The historical observations were converted into time-series data and analysed using the ARIMA model. The forecasting analysis generated future values based on patterns present in the historical observations. The forecast results were represented graphically to make the predicted trends easier to interpret.

Figure 1: Historical Weather Trends and ARIMA Forecast

The generated graph illustrates the historical observations together with the forecasted values. The visualization provides a clear representation of the temporal behaviour of the selected weather parameter and its projected future trend.

D. City-Wise Climate Analysis

The analysis also provided a comparison of weather patterns across Nagpur, Mumbai, Pune, Hyderabad, and Delhi. Differences in temperature, rainfall, humidity, wind speed, and atmospheric pressure were observed through the calculated values and visualizations. This city-wise analysis demonstrates that weather patterns can vary considerably across different geographical locations.

E. Discussion

The results indicate that combining exploratory analysis, correlation analysis, time-series analysis, and ARIMA forecasting provides a structured approach to weather-data analysis. The findings are consistent with the general use of statistical time-series models for identifying temporal patterns and forecasting future observations. However, the performance of ARIMA depends on the characteristics and quality of the historical dataset.

The study also shows that forecasting alone does not provide a complete understanding of weather behaviour. Combining forecasting with climate analytics and visualization provides additional information about historical trends, parameter relationships, and city-wise variations. Therefore, the proposed R-based approach can serve as a useful framework for analysing historical weather data and generating statistical forecasts.



V. CONCLUSION

This study developed a structured approach for **weather forecasting and climate analytics using R** based on the Kaggle *weather forecasting dataset*. The dataset consists of **3,655 observations from Nagpur, Mumbai, Pune, Hyderabad, and Delhi**, covering temperature, humidity, rainfall, wind speed, and atmospheric pressure. Using RStudio, the study organized and pre-processed the historical observations and applied exploratory data analysis, correlation analysis, and visualization to examine variations and relationships among these weather parameters.

The study further applied the **ARIMA time-series model** to historical weather observations for forecasting future values. The analysis demonstrates that combining historical weather analysis with statistical forecasting provides a more comprehensive view of weather behaviour than forecasting alone. The city-wise analysis also provides a basis for examining differences in weather patterns across the five selected locations.

The proposed framework can be extended by incorporating **real-time weather data, additional cities, and advanced forecasting techniques such as Random Forest, XG Boost, and LSTM**. Future work can also compare the forecasting performance of ARIMA with machine-learning and deep-learning models and integrate the analysis into an **R Shiny dashboard** for interactive weather monitoring and forecasting.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support and academic facilities provided by **Guru Nanak Institute of Technology, Nagpur**, which contributed to the completion of this study. The authors also acknowledge the use of the publicly available weather dataset obtained from **Kaggle**, which served as the primary data source for the analysis. The authors further extend their appreciation to the faculty members and peers whose guidance and constructive suggestions contributed to the development of this work

REFERENCES

- [1] Y. Lai and D. A. Dzombak, "Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation," *Weather and Forecasting*, vol. 35, no. 3, pp. 959–976, 2020, doi: 10.1175/WAF-D-19-0158.1.
- [2] K. K. Gill, K. Bhatt, B. Kaur, and S. S. Sandhu, "ARIMA Approach for Temperature and Rainfall Time Series Prediction in Punjab," *Journal of Agrometeorology*, vol. 25, no. 4, pp. 571–576, 2023, doi: 10.54386/jam.v25i4.2250.
- [3] S. Kumari and P. Muthulakshmi, "SARIMA Model: An Efficient Machine Learning Technique for Weather Forecasting," *Procedia Computer Science*, vol. 235, pp. 656–670, 2024, doi: 10.1016/j.procs.2024.04.064.
- [4] M. Elsaraiti and A. Merabet, "A Comparative Analysis of the ARIMA and LSTM Predictive Models and Their Effectiveness for Predicting Wind Speed," *Energies*, vol. 14, no. 20, p. 6782, 2021, doi: 10.3390/en14206782.
- [5] N. Mungale and J. Shinde, "Rainfall Forecasting: A Comparative Analysis of Deep Learning and Machine Learning Models with Application to Environmental Data," *International Journal of Intelligent Systems and Applications in Engineering*, 2024.

ABOUT THE AUTHORS

Sanika Sontake is a B. Tech. student in Computer Science Engineering at Guru Nanak Institute of Technology, Nagpur, affiliated with RTM Nagpur University. Her academic interests include data analytics, data science, statistical analysis, and data visualization. project work includes weather forecasting and climate analytics using R, involving exploratory data analysis, visualization, and statistical forecasting techniques. She is currently developing her skills in R programming, Python, SQL, and Power BI, with an interest in pursuing a career in data analytics and data science.

Oshini Jadhav is a B. Tech student in Computer Science and Engineering at Guru Nanak Institute of Technology, Nagpur, affiliated with RTMNU . She is a full-stack developer with experience in React.js, Next.js, and Java. Her areas of interest include web



development, data analysis, and statistical computing, with a focus on applying data-driven techniques such as R-based analysis and time-series forecasting to real-world problems.

Ishwari Waghule is a B. Tech student in Computer Science and Engineering at Guru Nanak Institute of Technology, Nagpur. she specializes in AI-driven data analytics development, pipelines, and Android mobile unsupervised learning applications for consumer behavioral datasets.

Prof. Roshni Kolhe is an Assistant Professor in the Department of Computer Science and Engineering at Guru Nanak Institute of Technology (GNIT), Nagpur. She holds B.E. and M. Tech. degrees. Her research interests include Data Science, Data Mining and Visualization, Digital Economy, and Cryptography & Network Security. She is actively involved in guiding student projects and fostering research-oriented learning.

Dr. Sudhir N. Shelke is the Principal and Director of Guru Nanak Institute of Technology (GNIT), Nagpur. He holds a Doctor of Engineering degree and possesses extensive expertise in VLSI design, microelectronics, and AI-driven systems. He is actively supporting our project by providing academic leadership and strategic guidance.