Hybrid SARIMA–LSTM Model for Local Weather Forecasting: A Residual-Learning Approach for Data-Driven Meteorological Prediction

Shreyas Rajeev
Stevens Institute of Technology
Hoboken, USA
srajeev@stevens.edu

Karthik Mudenahalli Ashoka
Stevens Institute of Technology
Hoboken, USA
kmudenah@stevens.edu

Amit Mallappa Tiparaddi
Stevens Institute of Technology
Hoboken, USA
aitiparad@stevens.edu

I. INTRODUCTION

Accurate weather prediction is a must-have for different areas, for example, agriculture, transport, disaster management, and energy planning. Committed to making correct predictions, they could greatly reduce losses, increase the safety, and optimize the use of the available resources. On the other hand, meteorological data are very complicated and always changing because they depend on many factors such as temperature, humidity, wind speed, and atmospheric pressure. On top of that, those data may also present both seasonal recurring patterns and abnormal changes which further complicate accurate forecasting.

The conventional statistical models like the Seasonal Autoregressive Integrated Moving Average (SARIMA) have been the primary instrument for time-series forecasting for many years due to their confidence, understandability, and the ability to depict linear and seasonal trends. Nevertheless, SARIMA models are incapable of solving problems related to nonlinear dependencies and sudden changes of meteorological patterns. On the other hand, deep-learning techniques such as Long Short-Term Memory (LSTM) networks can handle nonlinear relationships and temporal dependencies over the long run very well but they also need large datasets and substantial computational power.

This research puts forward a hybrid SARIMA–LSTM model to overcome these shortcomings which essentially combines the two methods' advantages. The SARIMA part of the model explains the data's linear and seasonal nature, and the LSTM part identifies the nonlinear residuals that SARIMA fail to model. Consequently, this merger expedites the accomplishment of higher prediction accuracy, stability, and generalization in weather forecasting applications.

The research data is a collection of various weather parameters recorded between January 2015 and August 2025 and is taken from the National Oceanic and Atmospheric Administration (NOAA). The instruments are assessed with the help of performance criteria such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), allowing for an extensive comparison. By applying such an approach, the research intent is to uncover the scenarios when each model yields the best results and to provide evidence that a combined SARIMA–LSTM framework can achieve better forecasting accuracy in practical weather prediction tasks.

II. RELATED WORK

Recently, time series forecasting models for weather and environmental data have been heavily studied, with much interest in hybrid frameworks combining statistical techniques with some deep learning methods. Models such as the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short Term Memory (LSTM) neural networks have now become a leading methodology for predicting time series data, largely due to their ability to accommodate temporal dependencies, seasonal effects, and nonlinearities often observed in empirical datasets. The application of these models have been conducted in areas ranging from predicting temperatures in buildings, to weather prediction, to rainfall nowcasting, to renewable energy estimating. In this section, we will review related literature that includes these models or their combinations, highlighting their methods, contributions and how they relate to the current study.

Cao, Li, and Sun [1] developed a comparative SARIMAX-LSTM model in predicting inner wall temperatures of residential building facades in Xi'an, China. The inner wall temperature prediction could be considered a way to find the "sweet spot" with respect to the energy efficiency of intelligent buildings, which also relates to the improvement in indoor comfort. The research by the authors compares a Seasonal AutoRegressive Integrated Moving Average (SARIMAX) model against a Long Short Term Memory neural network. The SARIMAX model performed reasonably well, with a MAPE of 0.24% and an IA of 0.98 meanwhile, the LSTM model had a MAPE of 1.45% and an IA of 0.94, correspondingly. Regression analysis was also done in that study. The most important predictors were the indoor temperature and the external wall temperature. Even though there were lower coefficients for solar radiation and outdoor temperature, the same predictors were indirectly responsible for the inner wall temperature. This research study is particularly relevant to the hybrid time series modeling, as it suggests a SARIMA type model may outperform deep learning methodologies when there are exogenous variables included, especially in case of a stable seasonal time series that has little nonlinearity. The paper certainly produced compelling evidence to suggest circumstances when SARIMA type models may not be relevant or perform equally or better than other machine learning models.

Kumari and Muthulakshmi [2] proposed a highly effective machine learning framework for weather forecasting based on SARIMA, which has been implemented on actual meteorological data sets (e.g., temperature, humidity, wind speed, and rainfall). This work fully adhered to a systematic time series modelling framework consisting of cleaning, differencing, normalization, and the selection of model order using the Akaike Information Criteria and Bayesian Information Criteria. The RMSE and MAE were 0.87 and 1.24 °C, respectively. The proposed framework was, therefore, suitable for short-term forecasting. The authors did emphasize that great care was taken in obtaining adequate quality input data and thus in the preprocessing stage. Specifically, noise reduction and stationarity are important practical steps to ensure that forecasts are more accurate. Additionally, although admittedly both works used SARIMA in different more general meteorological terms, but self similarly note that the simplicity of using SARIMA, low computational limitations, and performance under time consistent seasonality traits is a strong baseline modeling approach for forecasting any foregone environment. This paper, while noting the ability of deeper approaches to model and detect long term and non linear trends, still confirmed that SARIMA still remains relatively competitive for small to medium sized datasets, as that often occurs when interpretable and consistency are implicitly relevant.

He et al. [3] have advanced the field by developing a deep learning rainfall nowcasting model that combined LSTM and Decision Tree Regression. The primary inputs of the new model were GNSS derived from PWV and CAPE, along with hourly temperature, humidity, and wind variables sourced from 20 GNSS stations throughout Taiwan. The authors proposed their model as the Deep Learning Rainfall Nowcast (DLRN) model, and reported an RMSE of 1.25 mm, MAE of 0.37 mm, and correlation coefficient of 0.75, outperforming potentials of conventional machine learning, as well as a single deep learning technique, by 34%. More notably, the feature-importance analysis revealed that 43% of CAPE and 35% of PWV were the two most influential predictor contributors for rainfall occurrence. Overall, this hybrid modeling approach demonstrates the importance of applying deep learning and interpretability by integrating physically meaningful atmospheric predictors. In contrast to the purely statistical SARIMA models, which rely solely on recognizing time correlations, the proposed model provides predictive and interpretable value using physically meaningful variables along with a neural network learning component, which captures the dynamic meteorological process. Finally, the study connects traditional physics based modeling approaches with data driven or AI forecasting a clear methodological principle that is at the basis of hybrid SARIMA-LSTM future studies.

Hossain, Shams, and Ullah [4] provided a detailed summary of a comparative investigation of ARIMA, SARIMA, and LSTM models to help forecast renewable energy in Dhaka, Bangladesh. The authors obtained approximately ten years of hourly solar irradiance and wind energy data (2014-2023) from the NASA POWER database. The authors created and tested statistical and deep learning methods to forecast combined renewable energy density predictive analysis with a modeling time horizon of two years (2024-2025). Although their analyses showed that the ARIMA (2,1,2) and SARIMA (2,1,2)(1, 1, 1, 24) models captured some basic temporal structures, neither model was able to capture any nonlinear dynamics, as R^2 was close to zero for both models. Notably, the LSTM model vastly improved the models and showed excellent generalization via time series cross validation (mean R^2 was 0.9847) with the best model having $R^2 = 0.986$. During the analyses the LSTM displayed realistic seasonal patterns and stable longer term behavior which is further evidence of its ability to learn nonlinear dependencies and to preserve temporal continuity for renewable energy data. This should be a reassuring message to the research community and practitioners alike, as there continues to be a growing body of research ready to suggest that while SARIMA is still a good choice for linear periodic patterns, deep learning models are now the model of choice for heavily dynamic nonlinear landscapes such as forecasting renewable energy.

E. Krascsenits and A. Kiss [5] conducted a comparative study on ARIMA and LSTM models for temperature forecasting using time series data for two different climate regions: Budapest and Los Angeles . The authors found that while traditional statistical methods, such as ARIMA, provide good interpretability and robustness for univariate weather series, LSTM models are better at encapsulating complex temporal dependencies. They evaluated LSTM and ARIMA with standard performance metrics and found that LSTM predicted better in the short run than ARIMA, albeit at a greater computational and training costs. The authors conclude that deep-learning architectures can be effective complements to traditional models when climate data is nonlinear or exhibits regional variability.

Machado, Da Paixão Ataíde, and Borges [6] investigated Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) architectures for short-term weather forecasting, using real data from Brazil's three local meteorological stations located in Brasília, Florianópolis, and Manaus. Forecasting was characterized as a multivariate regression problem where the atmospheric parameters of temperature, humidity, pressure, wind speed, radiation, and rainfall were selected as input. The study trained and tested the models with a temporal dataset spanning years 2021 to 2023. BiLSTM was found to outperform LSTM at all meteorological stations, delivering the lowest global Root Mean Squared Error (RMSE) values (65.56 for Florianópolis, 115.31 for Brasília, and 100.84 for Manaus) overall. The authors argued that the BiLSTM architecture can extract more powerful temporal features by allowing information processing in both forwards and backwards directions, which enhances predictive accuracy in regional climate systems.

Peng Chen, Aichen Niu, Duanyang Liu, Wei Jiang, and Bin Ma [7] evaluated temperature time series forecasting by the SARIMA model for monthly mean temperatures in Nanjing, China, from 1951 to 2017. The authors analyzed data from sixty-six years, with the SARIMA $(1, 1, 1)(1, 0, 1)_{12}$ model model identified as the best; the model had the lowest Akaike Information Criterion (AIC = −2754.63). Validation of the model identified a low Mean Squared Error (MSE ≈ 0.84–0.94) with test data from 2015-2017, indicating reliable predictability over the long-term. The study indicated statistical methods remain powerful and suitable methods for meteorological forecasting when data is periodic and stationary characteristics.

These works collectively demonstrate a methodological evolution of time series modeling over time, from a pure statistical paradigm to deep learning and hybrid paradigms. Cao et al. [1] and Kumari and Muthulakshmi [2] illustrate that SARIMA and its variants are still useful for those systems that possess seasonality, stability, and have smaller datasets. On the other hand, He et al. [3] and Hossain et al. [4] have presented deep learning methods that outperform others in nonlinear and large-scale scenarios. A recurring theme throughout these works is that the integration of the interpretability of SARIMA with the adaptiveness of LSTM yields the most balanced forecasting framework that can model both linear and nonlinear dependencies effectively. In the problem domain of meteorological and environmental forecasting with which the present work is concerned, the literature reviewed has shown that the most promise is held by

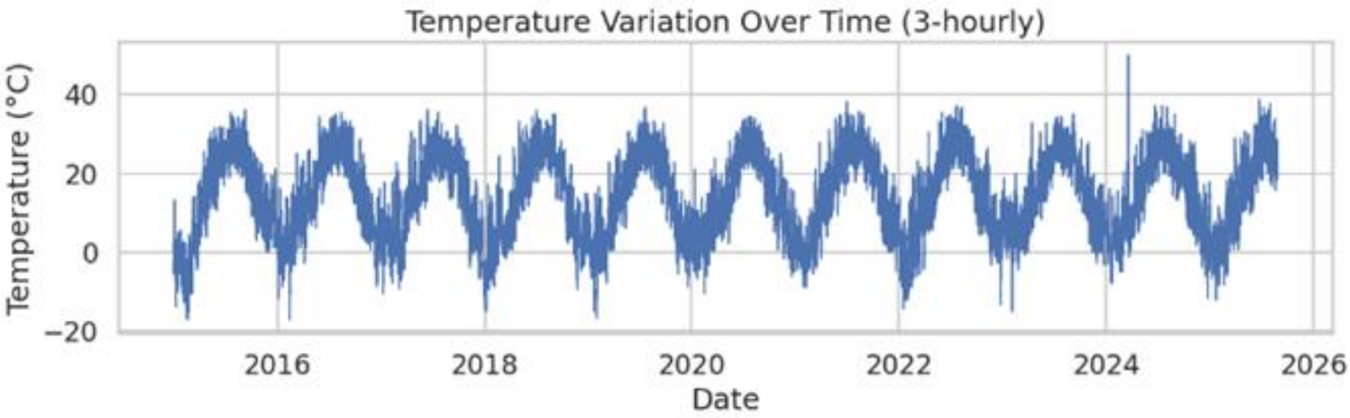

Figure 1 Temperature variation over 5 years

hybrid SARIMA-LSTM architectures. Consequently, SARIMA provides explainable, seasonally consistent modeling, while LSTM adds flexibility to capture transient and irregular patterns. Such a model can reduce forecast error, improve temporal stability, and support decision making on aspects like intelligent buildings, weather prediction, rainfall nowcasting, and renewable energy optimization. The collective findings from these studies thus set the theoretical and methodological foundation for the proposed research, which seeks to extend these hybrid approaches toward better predictive performance in short term environmental forecasting.

III. OUR SOLUTION

A. *Description of the Dataset:*

The data used in this research come from the local weather station at Newark Liberty International Airport (KEWR) that is part of the NOAA Integrated Surface Database (ISD). It records (Latitude: 40.68° N, Longitude: −74.17° W, Elevation: 2 m) . The file includes close to 129k observations at 3-hour intervals between 2015 and 2024, making it one of the most comprehensive local level climate records in the area. The data represent measurements of various meteorological parameters such as air temperature (TMP) Fig. 1, dew point (DEW) Fig. 2, sea-level pressure (SLP), wind direction and speed (WND), and visibility (VIS). The time resolution here allows for the development of various short-term (1–7 day) forecasting experiments.

The whole data cleaning strategy that was implemented involved several well-planned steps to change the raw data from the National Oceanic and Atmospheric Administration (NOAA) Integrated Surface Database (ISD) into a proper and user-friendly format. The ISD datasets record multiple attributes within single fields separated by commas (e.g., TMP = -0033,5 indicates −3.3 °C, or WND = 230,1,N,0031,1 indicates wind direction 230° and speed 3.1 m s^{-1}). In order to get the first numeric token of each coded field string-splitting methods were used to isolate it.

After that unit conversions were put in place to make all variables standard TMP and DEW were divided by 10 to get degrees Celsius (°C), SLP was divided by 10 to get hectopascals (hPa), and the fourth element of WND was taken out and divided by 10 to get wind speed in meters per second (m s^{-1}). Non-numeric or invalid tokens were replaced by NaN values to keep the data clean, and all fields were explicitly cast to floating-point type to allow mathematical operations, thus forming a structured numeric dataset with consistent units that is suitable for univariate and multivariate modeling.

B. *Machine Learning Algorithms:*

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is a conventional statistical method that is mainly used in time-series forecasting. It is a very powerful tool for data that exhibit both trend and seasonal patterns. SARIMA achieves this by adding to the ARIMA model the seasonal terms that represent the regular fluctuations over a certain time period such as the hourly or yearly temperature variations.

SARIMA is a very understandable and quick to compute model, which are factors that make it great for stable and periodical datasets. In weather forecasting, it is a practice that gives good results when temperature, humidity, or pressure are the characteristics that follow a certain seasonal cycle. The model gets rid of the trend by differencing and captures the temporal dependencies through autoregressive and moving average components.

On the other hand, SARIMA has the drawback of linearity between past and future observations, which hampers the capacity of the model to handle complex nonlinear behaviors that are frequently found in meteorological data. The model can only be used effectively in data series with obvious periodic patterns and relatively smooth transitions between them. Although it has such a limitation, SARIMA still represents a good starting point in uncovering linear trends and acts as a dependable baseline in this research.

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad (1)$$

where:

- B is the backshift operator
- Y(t) is the observed value at time t
- E(t) is the random error term

Long Short-Term Memory (LSTM) is a sophisticated deep learning model that is part of the Recurrent Neural Networks (RNNs) family. The model is targeted explicitly to identify dependencies and relationships in temporal data sequences. LSTM networks can recall and alter information over lengthy

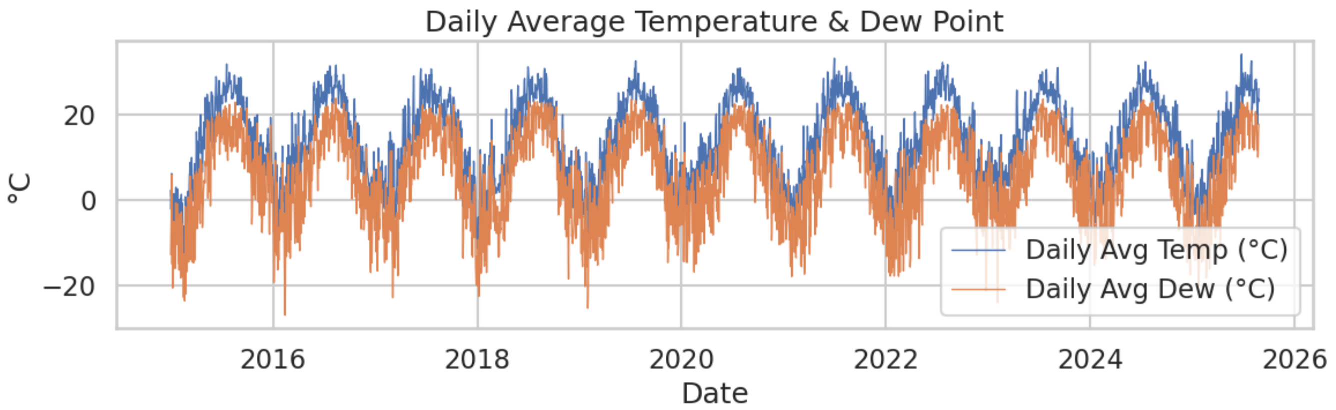

Figure 2 Daily average of temperature and dew

time periods, which is why they are the most powerful of models that can capture complex, nonlinear changes in time-series data.

The main strength of LSTM models is their capability to process non-stationary and nonlinear data, thus they are perfect for simulating atmospheric processes which are highly unpredictable and rapidly changing. Nevertheless, LSTM models usually have to be trained with more data and require more computational power than traditional methods.

In the case of weather forecasting, LSTM networks can discover intricate relationships between several meteorological variables such as temperature, humidity, wind speed, and atmospheric pressure. Unlike SARIMA that models relationships via fixed linear equations, LSTM learns patterns on the fly by being trained on the historical data. Therefore, it is capable of adjusting itself to sudden changes and irregularities in weather.

LSTM is the primary method that can deal with non-stationary and nonlinear data problems, thus it is the best choice to model atmospheric processes that have rapidly changing and unpredictable dynamics. A large number of data and a high computational power are also required in LSTM models.

$$
\begin{aligned}
f_t &= \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \\
i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\
\widetilde{C_t} &= \tan\mathrm{h}(W_c[h_{t-1}, x_t] + b_c) \\
C_t &= f_t * C_{t-1} + i_t * \widetilde{C_t} \\
o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\
h_t &= o_t * \tan\mathrm{h}(C_t)
\end{aligned} \tag{2}
$$

where x(t) is the input vector at time t, h(t) is the hidden state, C(t) is the cell memory, and σ and tanh denote activation functions.

Both SARIMA and LSTM models have a powerful prediction ability when used separately, but they each have certain types of limitations. For instance, SARIMA is great for capturing linear and seasonal factors but has difficulty with non-linear dependencies. On the other hand, LSTM can uncover nonlinear patterns effectively, though it may not clearly capture seasonality or trend structures.

In an attempt to overcome these weaknesses, the current research features a hybrid SARIMA – LSTM model that integrates the two strategies. Weather data's linear and seasonal behavior is modeled by SARIMA in this setting. The residual errors that depict the parts of the data in which SARIMA fails are then fed into the LSTM model. The LSTM gets these residual nonlinear patterns, and the final prediction is made by combining both models' outputs.

This combined methodology makes use of the merits of both methods — SARIMA's interpretability and ability to capture seasonality, and LSTM's flexibility in handling nonlinear and complex dependencies. Consequently, the hybrid SARIMA–LSTM model results in enhanced forecasting performance with lower error rates and greater consistency than either model alone.

C. Implementation Details

During the implementation stage, the SARIMA–LSTM hybrid forecasting model in Python was developed and experimentally evaluated with the use of the statsmodels and TensorFlow libraries. In order to capture the linear and seasonal patterns of the data, the SARIMA module was fitted to the temperature time series, while the residuals (random variations) were accounted for by a two-layer LSTM network with 64 and 32 hidden units, ReLU activation, and the Adam optimizer. In addition, each LSTM input sequence represented 30 time steps of residuals and some meteorological features such as the dew point, pressure, wind speed, and visibility.

The hybrid forecast was, therefore, the sum of the SARIMA prediction and the LSTM-estimated residual. The hybrid model, in fact, resulted in a significant performance improvement as it obtained the Mean Absolute Error (MAE) of 1.48 °C and the Root Mean Square Error (RMSE) of 1.98 °C, thus, it was better than the single SARIMA model (MAE 1.91 °C) and the LSTM model (MAE 1.72 °C). The present results serve as a positive signal that residual learning is an effective way to improve the accuracy of short-term temperature prediction and, at the same time, it is computationally efficient.

D. Novel Solution

In this project, we will be developing a hybrid forecasting model that leverages the advantages of SARIMA (Seasonal AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) neural networks. To illustrate, SARIMA models can capture linear trends and seasonal patterns in time-series data quite well, but they are not capable of handling non-linear dependencies. On the other hand, LSTM models can uncover complex, non-linear relationships;

however, they may not perform well in the case of strongly seasonal or trending data.

Understanding these drawbacks, our solution combines both methods in a two-stage hybrid model:

- First stage - SARIMA Component: The SARIMA is initially used to fit and forecast the linear and seasonal parts of the data.
- Second stage - LSTM Component: The leftover errors (actual minus SARIMA-fitted values) are then captured through an LSTM network.

The forecasts from the two models are finally merged to obtain the final prediction. Such a hybridization is a novel way of harnessing both statistical and deep learning paradigms to enhance prediction accuracy and stability. It enhances forecast precision compared to standalone traditional or neural models.

REFERENCES

[1] [W. Cao, H. Li, and Y. Sun, "Using SARIMA and LSTM Models to Forecast the Temperature of Internal Walls of Building Facades: A Case Study of Residential Buildings in Xi'an," in Proc. 37th Chinese Control and Decision Conf. (CCDC), Xiamen, China, 2025, pp. 1635–1640. doi: 10.1109/CCDC65474.2025.11090431.

[2] S. Kumari and P. Muthulakshmi, "SARIMA Model: An Efficient Machine Learning Tool for Weather Forecasting," Procedia Computer Science, vol. 235, pp. 656–670, 2024.

[3] L. He, Y. Zhang, Q. Li, J. Zhao, H. Wang, and T. Zhang, "Deep Learning-based Feature Importance for Rainfall Nowcast based on GNSS PWV and CAPE," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 18 no. 1, pp. 26688–26698, 2025. doi: 10.1109/JSTARS.2025.3621857.

[4] M. L. Hossain, S. M. N. Shams, and S. M. Ullah, "Time-Series and Deep Learning Techniques for Renewable Energy Forecasting in Dhaka: A Comparison among three models (ARIMA, SARIMA and LSTM)," Discover Sustainability, vol. 6, art. no. 775, 2025.

[5] E. Krascsenits and A. Kiss, "Comparative Analysis of ARIMA and LSTM Models for Weather Forecasting Using Time Series Data," *Proc. 23rd Int. Symp. Intell. Syst. Informatics (SISY)*, Subotica, Serbia, pp. 119–124, 2025, doi: 10.1109/SISY67000.2025.11205407.

[6] C. E. Machado, K. R. Da Paixão Ataíde, and V. R. P. Borges, "Long Short-Term Memory Approaches for Weather Forecasting from Local Stations," *Proc. 9th Int. Conf. Frontiers Signal Processing (ICFSP)*, Paris, France, pp. 123–127, 2024, doi: 10.1109/ICFSP62546.2024.10785333.

[7] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time Series Forecasting of Temperatures Using SARIMA: An Example from Nanjing," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 394, no. 5, p. 052024, 2018, doi: 10.1088/1757-899X/394/5/052024.