International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/

SCHOLAR AI
Be Smart

Vol. 50 No. 1 April 2025, pp. 69-96

Hybrid LSTM-RNN And Sarima Modeling For Time Series Temperature Prediction: The Case Of Antananarivo, Madagascar

RALIJAONA Herilala Tafitasoloniaina[1], RANDRIANJA Malanto Miangaly[2], RANDRIANJA Roger[3]

[1] Informatique Department, Water and Environmental Engineering Department, University of Itasy,
Soavinandriana, Madagascar
tafitaralijaona@gmail.com
[2] Informatique Department, University of Itasy,
Soavinandriana, Madagascar
miangalirandrianja@gmail.com
[3] Water and Environmental Engineering Department, University of Itasy,
Soavinandriana, Madagascar
rgr.rand@gmail.com

Abstract: This study evaluates various approaches for temperature forecasting in Antananarivo, Madagascar, by comparing the performance of LSTM and SARIMA models, as well as their hybrid combinations. One of the explored strategies involves using an LSTM model to generate an initial forecast, and then modeling its residuals with SARIMA to refine the results. Another approach relies on using SARIMA to produce a preliminary estimate, whose predictions are subsequently incorporated into an LSTM model to better capture the complex dynamics of temperature variations. The goal is to identify the method that offers the best accuracy and stability to improve the reliability of weather forecasts.

While SARIMA has proven effective for linear data, it struggles to capture the non-linear fluctuations of local temperatures. The LSTM model, with its ability to model long-term dependencies and non-linearities, aims to address these limitations.

Our study demonstrated that combining forecasting models, particularly the SARIMA_LSTM hybrid approach, offers superior accuracy compared to other models for temperature forecasting in Antananarivo. This approach, which integrates exogenous variables into the SARIMA model and uses LSTM to refine forecasts by modeling residuals, consistently produced the most accurate results.

The performance of the models was rigorously evaluated using several statistical metrics and tests. We used the Root Mean Square Error (RMSE) to measure forecast accuracy, and Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to analyze the temporal structure of the data and residuals. Additionally, we applied the Ljung-Box test to verify the absence of residual autocorrelation, and analyzed skewness and kurtosis to understand the distribution of residuals. Finally, the heteroscedasticity test was performed to evaluate the constancy of error variance.

These analyses confirmed that the SARIMA_LSTM model, by leveraging the strengths of each constituent model, successfully captured both seasonal trends and complex relationships in the temperature data, leading to more reliable forecasts.

Keywords: Modeling, hybrids, prediction, LSTM, SARIMA, temperature, Antananarivo, Madagascar, Neural Networks, seasonality.

I. INTRODUCTION

Weather forecasting, particularly long-term forecasting, is of crucial importance to many sectors, from agriculture and energy to public health and natural disaster management. In Antananarivo, the capital of Madagascar, this forecasting is all the more essential as the city faces complex environmental challenges related to global climate change and increasing pollution.

Climate change, characterized by more frequent and intense extreme weather events, has direct consequences on temperature regimes. Prolonged drought episodes, alternating with periods of torrential rain, disrupt natural cycles and jeopardize the livelihoods of populations. Furthermore, Antananarivo's rapid urban growth is accompanied by an increase in atmospheric pollution, notably due to vehicle and industrial emissions. This pollution can modify the thermodynamic properties of the atmosphere and thus influence local temperature regimes.

To improve the accuracy of forecasts, we explored the use of recurrent neural networks, notably the LSTM (Long Short-Term Memory) model, recognized for its ability to model long-term temporal dependencies. To further optimize predictions, we implemented two hybrid approaches: LSTM_SARIMA, where the residuals of the LSTM model are modeled by SARIMA, and SARIMA_LSTM, where the predictions of the SARIMA model are used as input for the LSTM.

This study compares the performance of SARIMA, LSTM, and their hybrid approaches to evaluate their effectiveness in forecasting temperatures in Antananarivo. The objective is to identify the most suitable method based on local climatic particularities and prediction accuracy.

II. STUDY DATA DESCRIPTION

A. Temperature data

The temperature data we have covers the period from early 2020 to August 31, 2024, representing a four-year time series. The data is recorded daily, as indicated by the density of points on the graph. The temperature, measured in degrees Celsius (°C) on the vertical axis, ranges from 12°C to 26°C throughout the observed period.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

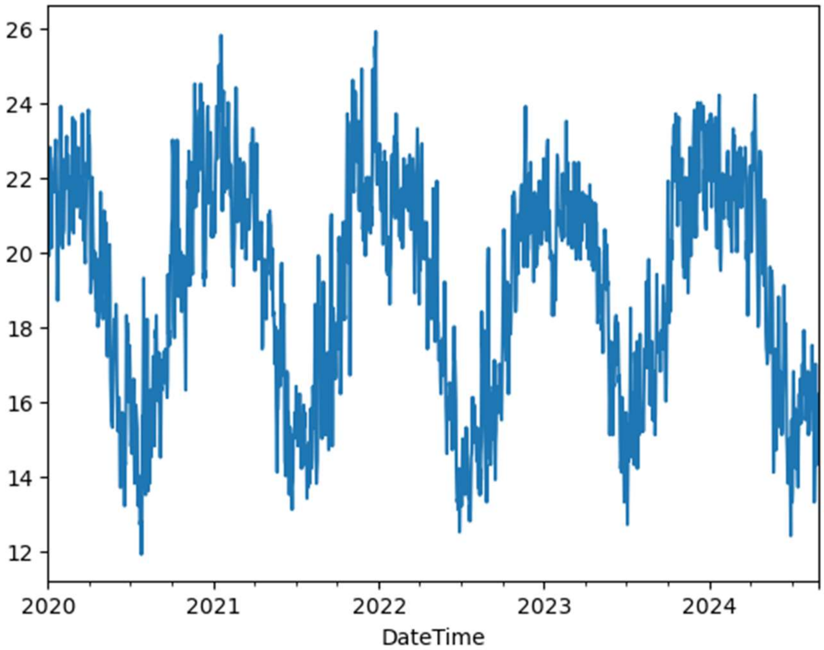


Figure 1. Température Data plot

Observations and Key Characteristics

The data reveal a strong seasonality, marked by regular annual cycles. Each year, we observe high temperature peaks followed by significant drops, indicating a climate with distinct hot and cold seasons. The temperature peaks appear to occur at roughly the same time each year, thus confirming an annual periodicity. The amplitude of temperature variations (the difference between peaks and troughs) seems stable from year to year, although slight variations can be observed. In addition to this seasonality, short-term fluctuations are visible, indicating daily or weekly temperature variations. Over the entire study period, no clear long-term trend (neither upward nor downward) is evident in the average temperatures, although further analysis is needed to confirm this absence of trend. Finally, the data do not show any notable outliers, suggesting that the data are relatively clean and consistent.

B. Precipitation data

The data covers the same period as the temperature data, from the beginning of 2020 to August 31, 2024. They are recorded at a daily frequency, similar to the temperature data. The range of precipitation values varies considerably, from 0 mm to values exceeding 90 mm during significant peaks.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

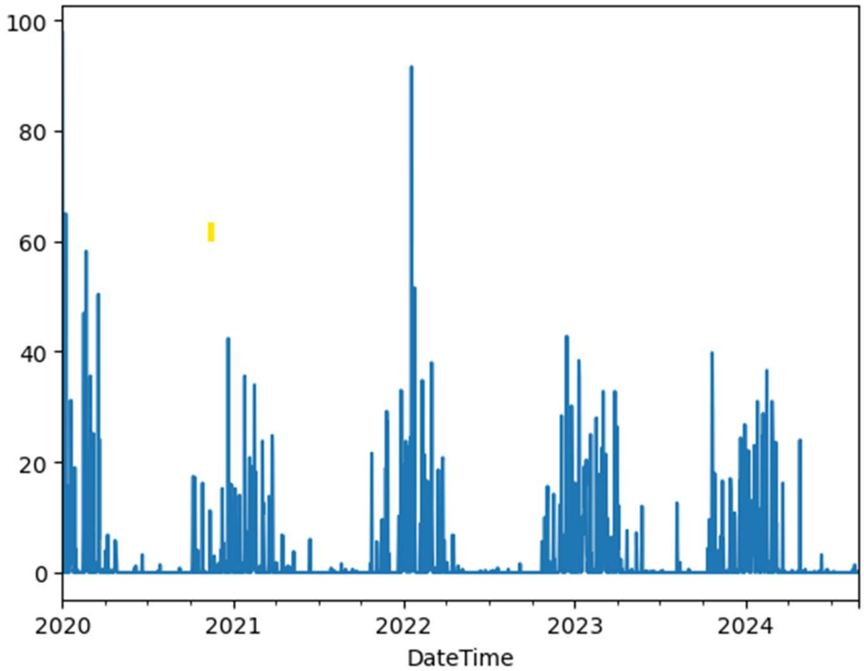


Figure 2. Precipitation Data plot

Observations and Key Characteristics :

Precipitation is highly sporadic, with long periods of drought interspersed with intense rainfall peaks. Several particularly high precipitation peaks are observed, suggesting heavy rain or storm events. The variability of precipitation is high, with abrupt and significant changes over short periods. The clustering of precipitation peaks at certain times of the year indicates seasonality in this variable. There is no clear long-term trend in precipitation, just as there is no clear long-term trend in temperature. Finally, many data points are close to zero, indicating prolonged periods without precipitation.

C. Wind speed data

In our study data, wind speed oscillates between approximately 0 m/s and slightly over 10 m/s. We included this parameter because it influences temperature both directly and indirectly. Directly, it causes wind chill, mitigates temperature variations through air mixing, and increases evaporation, thereby cooling the surface. Indirectly, it affects cloud formation and movement, influencing sunlight and precipitation, and alters air humidity.

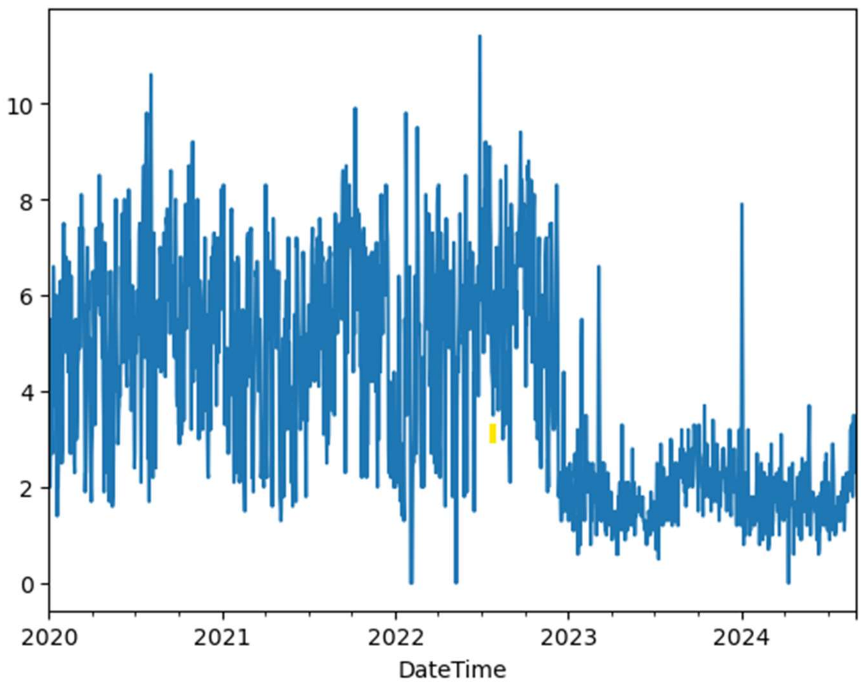


Figure 3. Wind speed Data plot

Observations and Key Characteristics :

Wind speed exhibits notable variability, with frequent fluctuations throughout the analyzed period, though it does not show a clear seasonality, unlike temperature. While no obvious seasonal pattern is observed, there may be subtle seasonal variations requiring further analysis. The time series is also characterized by regular peaks and troughs, suggesting sudden changes in wind dynamics. A regime shift is apparent at the beginning of 2023, with a significant reduction in variability and lower average values, which could reflect a change in local weather conditions or a data collection issue. Finally, no long-term trend, either upward or downward, is clearly identifiable in the wind speed data over the entire study period.

III. DEVELOPMENT AND IMPLEMENTATION OF FORECASTING MODELS

A. SARIMAX Model

Initially, a Seasonal ARIMA (SARIMA) model was fitted to predict temperature based on time series data, incorporating two exogenous variables: rainfall and wind speed. The dataset was split into training (80%) and testing (20%) sets. The model was then optimized by automatically selecting the best parameters using auto_arima, considering seasonality and exogenous variables. Once the optimal parameters were identified, temperature values were predicted on the test set.

B. LSTM Model

Subsequently, an LSTM model was used to predict temperature using historical meteorological data, including temperature, precipitation, and wind speed, with the addition of month and day as seasonal variables. The data is loaded from a CSV file, with the creation of a DateTime column to index observations by date. The variables to be predicted (temperature) and explanatory variables (precipitation, wind speed) are converted to numerical format, and a fixed-size sliding window of '10 days' is used to create the model's input sequences, where each window includes the temperatures and precipitation of the previous days. The targets correspond to the temperature of the following day.

The data is then divided into three sets: training, validation, and testing. Seasonal variables, such as month and day of the week, are added to the data as additional features to enrich the inputs. These variables are transformed to match the dimensions of the input windows. The LSTM model is configured with a simple architecture: an LSTM layer with 64 neurons, followed by dense layers to capture the complex relationships between the variables.

To improve the LSTM model's ability to predict temperature, seasonal variables were integrated. These variables, representing the month of the year and the day of the week, provide the model with crucial temporal information to understand the seasonal patterns present in the temperature data. To do this, the month and day of the week values were extracted from the time index of the dataset and then organized into a seasonal feature matrix. This matrix was then expanded to match the temporal structure of the LSTM input data windows, and finally added to the training, validation, and testing datasets. This approach will also allow the LSTM model to better understand the cyclic trends of temperature.

Next, the model is trained with an MSE loss function and the Adam optimizer, for 200 epochs with a dynamic learning rate using Learning Rate Schedule. The best model is saved to avoid overfitting.

After training, the model is tested on the training, validation, and testing sets. The results are compared to the actual values and visualized as graphs to evaluate the quality of the predictions. Checks are also performed to avoid the presence of missing values in the input data.

C. Hybrid Model

1) LSTM_SARIMA

This architecture uses an LSTM layer with 64 units to model temporal trends and a Dense layer with 8 units and ReLU activation to learn non-linear relationships. The final Dense layer, with a single unit and linear activation, predicts the temperature. The model is trained using the Adam optimizer, with a learning rate scheduled to dynamically adjust the learning rate during training. The loss function is the Mean Squared Error (MSE), ideal for regression tasks, and Early Stopping is used to halt training in case of overfitting. The best model is saved using ModelCheckpoint. After training the model, we make predictions on the training, validation, and test sets, and then calculate the residuals (difference between actual and predicted values). These residuals are then modeled with a SARIMA model to capture any trends or seasonality not detected by the LSTM. The final predictions are obtained by adding the SARIMA model's forecasts to the LSTM model's predictions, which corrects errors and improves model accuracy. This combined process leverages the strengths of both models, LSTM for capturing complex temporal dependencies and SARIMA for modeling seasonal residuals or other non-linear trends.

2) SARIMA_LSTM

This hybrid approach includes both a SARIMA model and an LSTM model. First, a model is set up by integrating exogenous variables such as rain and wind speed into a DataFrame. The auto_arima function is used to automatically find the best SARIMA parameters adapted to the training data, and then predictions on the test set are made. Subsequently, the residuals of the SARIMA model are extracted. These residuals are used to train an LSTM model. The LSTM is designed to predict the residuals, and its predictions are added to the SARIMA predictions to correct the latter.

IV. RESULTS AND DISCUSSION

The evaluation of the models' performance will be exclusively focused on the test dataset, thus ensuring an objective measure of their ability to generalize. The metrics used for this evaluation include:

- RMSE (Root Mean Squared Error)
- Ljung-Box test

- ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function)
- Heteroscedasticity test
- Skewness and Kurtosis

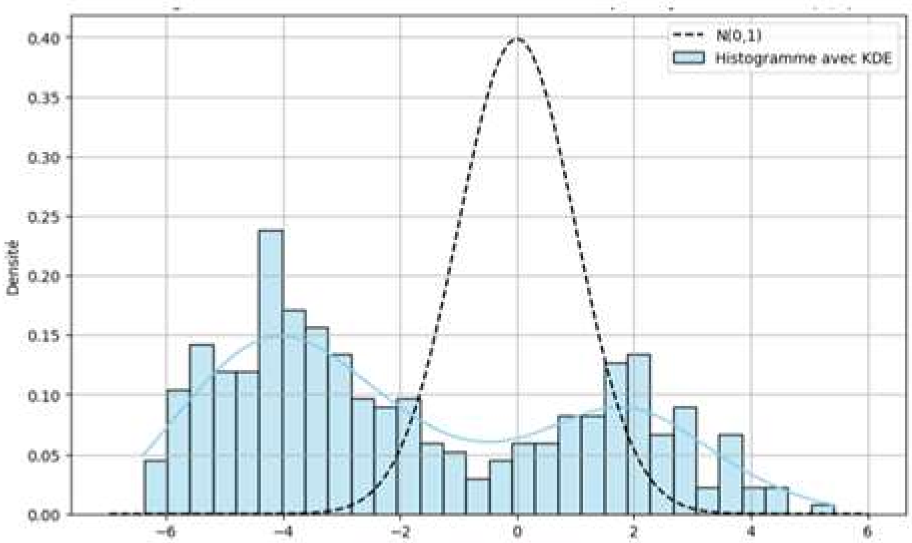
This approach ensures that the reported performances reflect the models' actual ability to make accurate predictions on data not seen during training.

A. SARIMAX

1) RMSE

The Root Mean Squared Error (RMSE) of the predictions on the test data is 3.482 °C for the SARIMAX model, indicating that, on average, the temperature predictions deviate from the actual values by 3.482 degrees Celsius. To assess the relevance of this value, it is essential to consider the scale of the observed temperatures. If the temperatures generally range between -10 °C and 40 °C, an RMSE of 3.482 °C represents a relatively small proportion of this range, suggesting that the model offers reasonable accuracy. However, for the case of Antananarivo, where temperatures fluctuate within a narrower range, between 12 °C and 27 °C, an RMSE of 3.482 °C is proportionally higher, indicating a less satisfactory performance.

2) Visual evaluations of residuals from Test Data



International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

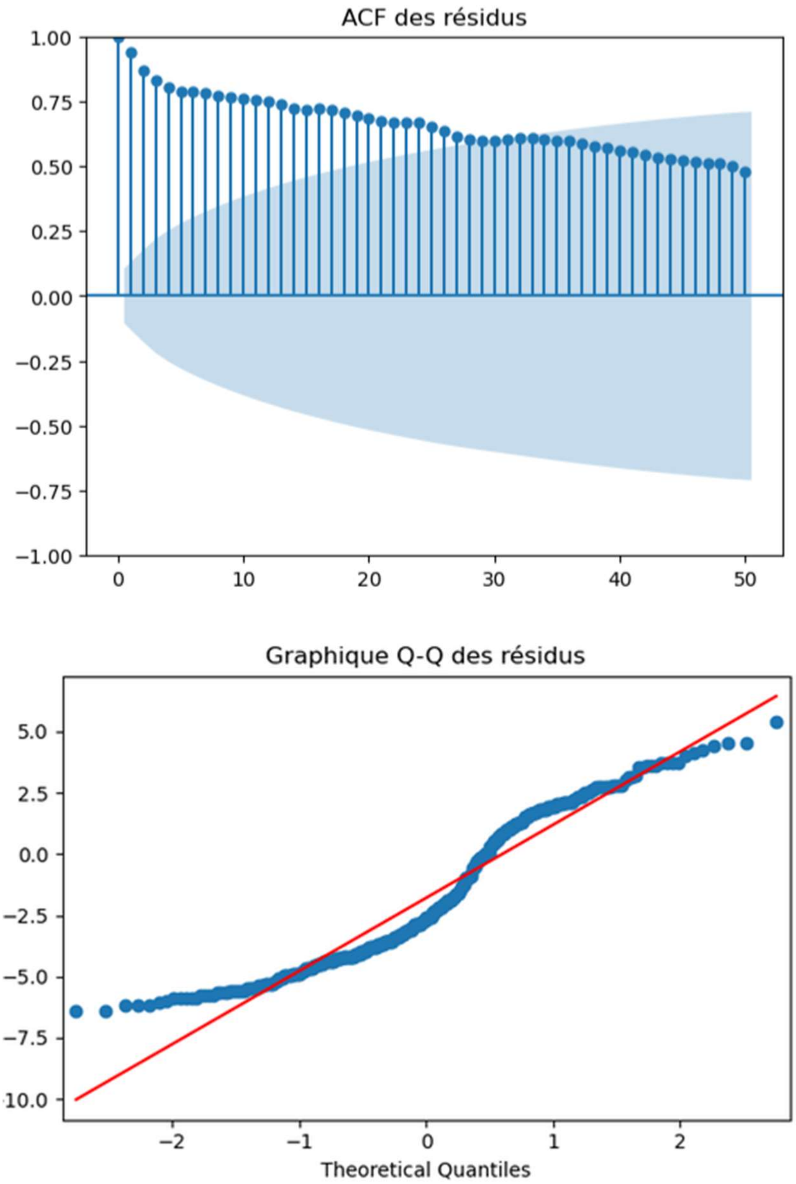


Figure 4. SARIMA Model Performance Evaluation: Graphical Analysis of Residuals

The graphical analysis of the SARIMA model's residuals on the test data reveals several key points. The residuals oscillate around zero, which is encouraging, but their volatility and amplitude variations over time, notably an increase in variance towards the end of the period, suggest heteroscedasticity. Additionally, the presence of an outlier may indicate specific anomalies in the data.

The histogram of the residuals shows a slightly skewed distribution with thicker tails than those of a normal distribution, which challenges the assumption of normality of the residuals. The autocorrelation function (ACF) of the residuals shows significant peaks outside the confidence interval at the first lags, as well as a slow decay thereafter, indicating residual autocorrelation and suggesting that the model has not fully captured the temporal dependencies of the data.

These observations highlight certain limitations of the SARIMA model, particularly regarding heteroscedasticity, non-normality, and autocorrelation of the residuals.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.
© 2025 Scholar AI LLC.
https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

3) Statistical evaluation of residual properties

Table 1. Table of residual analysis results

Ljung-Box (L10)(Q):	2292,843733
Prob(Q):	$1{,}54 \times 10^{-67}$
Heteroskedasticity (H):	191,05
Prob(H) (two-sided):	$1{,}87 \times 10^{-43}$
Skew:	0.415
Kurtosis:	-1,119

- Test de Ljung-Box (L10) :

The Ljung-Box test evaluates the null hypothesis that the data does not exhibit autocorrelation up to a certain number of lags. In our case, the test was performed with a lag of 10 (L10). A high test statistic and a very low p-value (1.536046e-67) indicate that the null hypothesis is rejected, suggesting significant autocorrelation of the residuals at this lag.

- Skewness and Kurtosis :

Skewness: A value of 0.41541490253907115 indicates a slight positive skewness of the residuals, meaning that the right tail of the distribution is longer or thicker than the left tail.

Kurtosis: A value of -1.1190581230822265 indicates that the distribution of the residuals is less peaked (more flattened) than the normal distribution, which is characterized by a kurtosis of 3.

- Breusch-Pagan Heteroscedasticity Test :

This test evaluates whether the variance of the residuals is constant (homoscedasticity). An extremely low p-value (1.8727783338092684e-43) suggests that the null hypothesis of homoscedasticity is rejected, indicating the presence of heteroscedasticity, i.e., the variance of the residuals varies as a function of the values of the explanatory variables.

The results obtained reveal that the SARIMAX model has limitations. It fails to fully capture temporal dependencies, which compromises the accuracy of predictions. Furthermore, the residuals deviate from normality, which can affect the reliability of statistical tests and confidence intervals. The integration of additional explanatory variables or the exploration of alternative models offering a better ability to capture the specificities of the data would constitute relevant approaches to improve the accuracy of forecasts. However, in the context of our study, we primarily focus on the comparative analysis of the models.

B. LSTM

1) RMSE

In the case of the LSTM model, we have an RMSE of 1.1276 degrees Celsius on the test dataset, which means that, on average, the prediction error of our model compared to the actual temperatures is approximately 1.1276 degrees. This value represents the standard deviation of the prediction errors and provides a measure of the overall accuracy of the model.

To assess the relevance of this result, it is essential to consider the specific context of the application and the characteristics of the data. In Antananarivo, where temperatures generally range between 12 and 27 degrees Celsius, an average deviation of 1.1276 degrees can be interpreted as relatively acceptable. Indeed, this variation represents a reasonable fraction of the typical temperature range, especially when considering the significant daily and seasonal fluctuations that can occur.

Although the RMSE of 1.1276 provides a valuable indication of the model's accuracy, it should not be interpreted in isolation.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

2) Graphical analysis of residuals

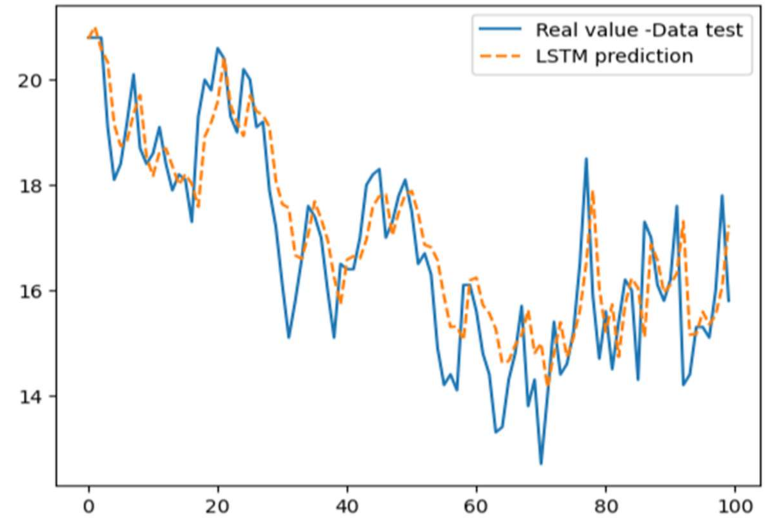


Figure 5. Difference between Actual Value and Value Predicted by the LSTM Model

The graph presents two curves: a blue curve representing the actual values of the test data, and an orange dotted curve representing the predictions of the LSTM model. The blue curve reveals a time series characterized by marked fluctuations and trends, with peaks and troughs that demonstrate some volatility. The orange curve, on the other hand, closely follows the actual values curve, although it exhibits some deviations. It is observed that the LSTM model manages to capture the general trends of the data, but it may struggle to reproduce certain details or peaks. A comparison of the two curves reveals an overall similarity, suggesting that the LSTM model is able to grasp part of the data structure. However, deviations are visible, particularly at points of rapid change or at sharp peaks and troughs. It seems that the model has a tendency to smooth out fluctuations, which can lead to less accurate predictions during abrupt changes.

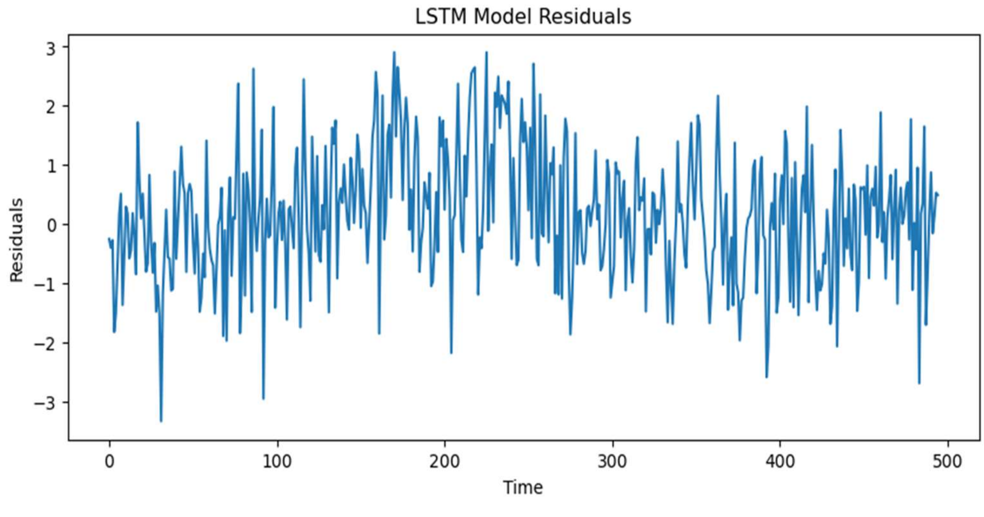


Figure 6. LSTM Model Residuals with Test Data

Figure 6 illustrates the evolution of the LSTM model's residuals over time, representing the prediction errors. The analysis reveals several significant elements. First, the residuals appear to be globally centered around zero, indicating the absence of systematic bias in the predictions. Second, their variability remains relatively constant, suggesting homogeneous accuracy across the entire dataset. Moreover, no marked trend or repetitive pattern appears, which demonstrates a random behavior of the errors. However, a few outliers stand out, reflecting moments when the model encountered difficulties in predicting accurately.

These observations confirm that the LSTM model adapts well to the data. The absence of an identifiable pattern in the residuals reinforces the idea of a correct fit, although the analysis of outliers could help identify potential improvements.

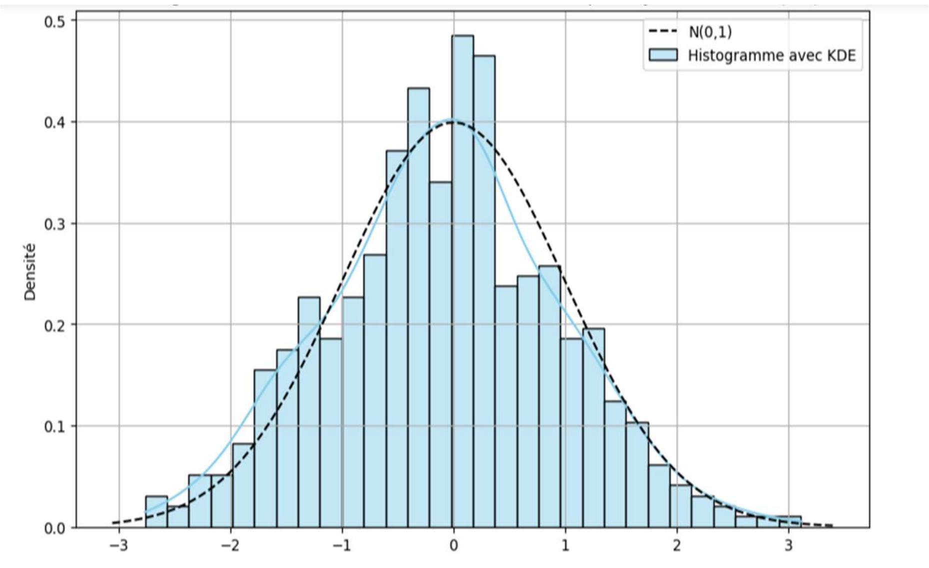


Figure 7. LSTM Residual Histogram with Kernel Density Estimation and N(0,1) Curve

General Observation:

The histogram reveals the distribution of the LSTM model's residuals, centered around zero and exhibiting a bell shape, which suggests a distribution close to normal. A slight skewness is noticeable, with a slightly longer tail on the positive side. The kernel density estimation (KDE) provides a smoothed representation of the residual distribution, confirming the bell shape and slight skewness. The standard normal distribution curve (N(0,1)) serves as a reference, and it is observed that the residual distribution (KDE) generally follows it, although with some deviations. The tails of the residual distribution appear to be slightly thicker than those of N(0,1), indicating a higher probability of observing extreme values.

Interpretation :

The analysis of the normality of the residuals reveals that they generally follow a distribution close to normal, but with some deviations. The slight skewness and thicker tails indicate that the distribution is not perfectly normal, which may have implications for the validity of statistical tests and confidence intervals. The fact that the residuals are centered around zero and generally follow a normal distribution suggests that the LSTM model is performing well. However, the deviations from normality indicate that there is room for improvement. The presence of an outlier may also signal problems during the learning phase.

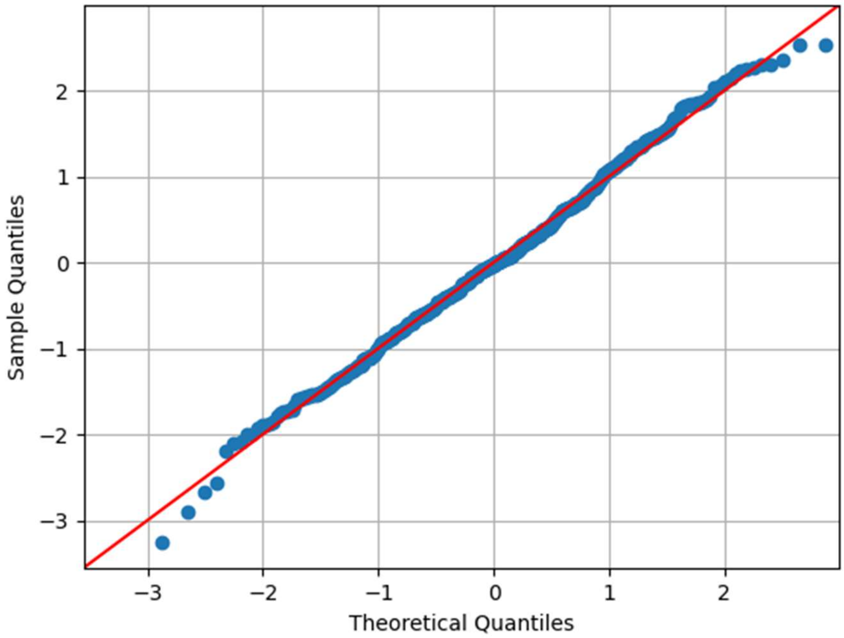


Figure 8. Graphique Q-Q des résidus – LSTM simple

General observation:

The graph presents the quantiles of the observed residuals, compared to the theoretical quantiles of a normal distribution. A red line serves as a reference, representing perfect alignment for a normal distribution. It is observed that the points are closely aligned with this line, with only a few slight deviations visible in the tails of the distribution. The majority of the points are located very close to the line, indicating a strong similarity between the distribution of the residuals and a normal distribution.

Interpretation :

The close alignment of the points on the red line suggests that the residuals of the LSTM model follow a distribution very close to normal. The slight deviations in the tails may indicate a slight deviation from perfect normality, but they are generally considered acceptable. This means that the statistical assumptions based on the normality of the residuals are likely valid for this LSTM model. The normality of the residuals is a positive indicator of the LSTM model's performance, suggesting that it is well-specified and captures the characteristics of the data well. Residuals that follow a normal distribution indicate that the model has successfully removed the systematic structures from the data, leaving behind random errors.

3) Statistical Analysis

- Autocorrelation Analysis

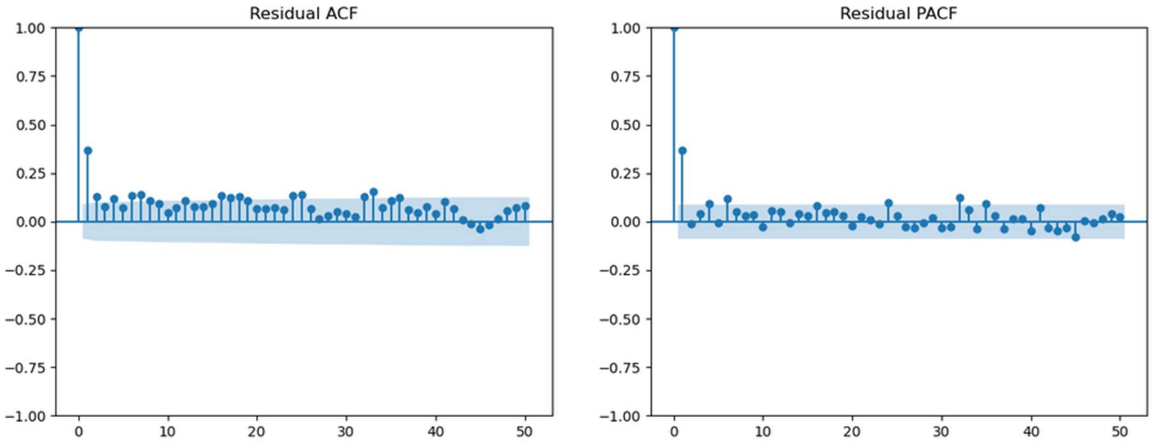


Figure 9. Courbe ACF et PACF for LSTM residuals

The ACF reveals several interesting aspects about the autocorrelation of the residuals. First, the high initial peak is expected, as the signal is perfectly correlated with itself at a zero lag. A noticeable peak at a lag of 1 suggests a marked correlation between residuals of consecutive time steps. Subsequently, a rapid decay of the ACF is observed, with most lags located within the confidence zone, indicating a rapid decrease in autocorrelation. However, a few lags slightly exceed this zone, which may reveal residual autocorrelation at certain time scales.

These observations have implications for the LSTM model. Significant autocorrelation at a lag of 1 suggests that some temporal dependencies have not been fully captured. The model could therefore be improved to better understand these short-term relationships. Despite an overall absence of correlation between the majority of the residuals, the presence of slight autocorrelation suggests potential for improvement. Adjusting hyperparameters, introducing new explanatory variables, or exploring more complex architectures could be considered to refine the model's performance.

The PACF curve highlights the direct relationships between residuals at different lags. As expected, a high initial peak appears at a zero lag, while a second noticeable peak at a lag of 1 reveals a marked partial autocorrelation between residuals of consecutive time steps. Subsequently, the decay is rapid, and most lags remain within the confidence zone. However, a few values slightly exceed this limit, suggesting the presence of residual partial autocorrelation at certain specific lags.

For the LSTM model, these observations indicate that some direct temporal dependencies have not been fully captured, particularly in the short term.

- Ljung-Box

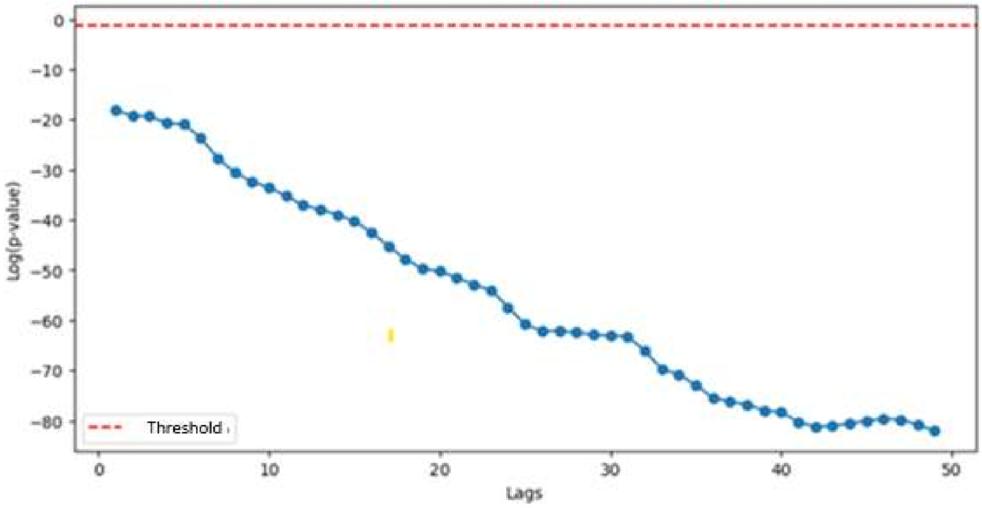


Figure 10. Log(p-value) Curves - Simple LSTM

The red horizontal line represents the 5% significance threshold. In statistics, when a p-value is less than 0.05 (or when the log of the p-value is less than log(0.05)), it means that the null hypothesis is rejected.

The blue curve, which shows the evolution of the log of the p-value as a function of the lags, remains largely below this threshold for all lags. The closer this curve is to zero, the smaller the p-value, which reinforces the idea of significant evidence against the null hypothesis. The null hypothesis of the Ljung-Box test states that the residuals are independent, i.e., there is no autocorrelation.

Since the blue curve remains well below the 5% threshold across all lags, we reject the null hypothesis. This means that there is significant autocorrelation in the residuals of your LSTM model. The results of the Ljung-Box test support the relevance of the hybrid approach.

- Skewness and kurtosis

Skewness : 0.057472137776929506

Interpretation :

In our case, the skewness is 0.057, which is very close to 0. This means that the distribution of the residuals is almost symmetrical, with a slight positive skewness (slightly longer tail on the right). This slight skewness is generally considered acceptable and does not indicate a major problem with the model.

Kurtosis: -0.11268350504713132

Interpretation :

The kurtosis is -0.113, which is less than 3. This means that the distribution of the residuals has slightly lighter tails than those of a normal distribution. This slight platykurtosis is also generally considered acceptable and does not indicate a major problem with the model.

The skewness and kurtosis values therefore indicate that the residual distribution of the LSTM model is very close to a normal distribution. The slight skewness and slight platykurtosis are generally considered acceptable and do not indicate a major problem with the model. These results reinforce confidence in the validity of the LSTM model's results and forecasts.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

- Breusch-Pagan Test :

Test Statistics : 3.3326

P-value : 0.0679

In this case, the p-value (0.0679) is greater than 0.05. This means that we cannot reject the null hypothesis. In other words, there is not enough statistical evidence to conclude that the variance of the LSTM model's residuals is not constant.

This result is positive, as it suggests that the variance of the LSTM model's errors is relatively stable. This reinforces the reliability of the model's forecasts, as homoscedasticity is a desirable condition for regression models.

C. LSTM_SARIMA

1) RMSE

For the LSTM_SARIMA hybrid model, the RMSE on the test dataset is 1.0356 °C. This means that, on average, the prediction error of this model compared to the actual temperatures is approximately 1.0356 °C, indicating a slight improvement over the simple LSTM model.

This variation represents a reasonable fraction of the typical temperature range, especially when considering daily and seasonal fluctuations. Although the RMSE is acceptable, its interpretation should not be done in isolation when evaluating the model's accuracy.

2) Graphical analysis of residuals

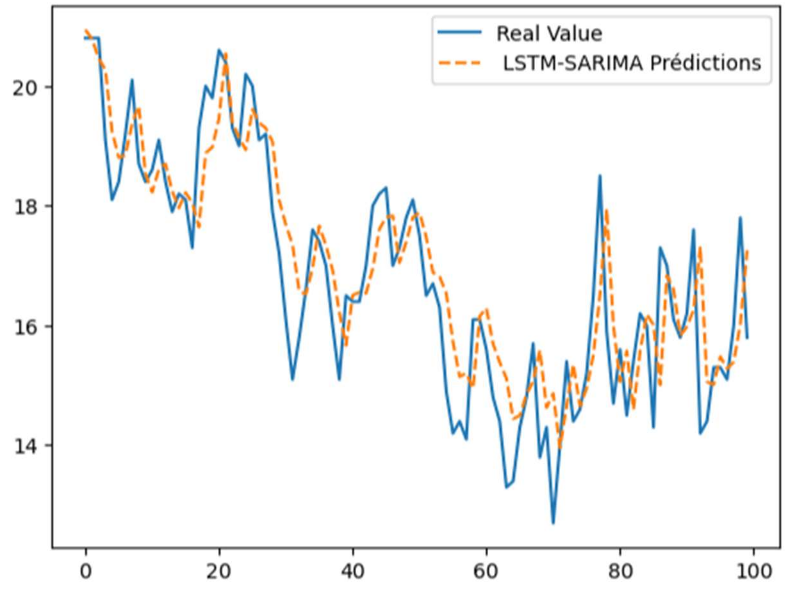


Figure 11. Gap Between Prediction and Actual Values

General observation:

The curve representing the predictions of the LSTM_SARIMA hybrid model closely follows the actual values, although some deviations remain. The model manages to capture the general trends of the data, but may miss certain specific details or peaks.

Comparing the two curves, an overall similarity is observed, indicating that the LSTM_SARIMA model is able to grasp part of the data structure. However, divergences appear, particularly during rapid changes or abrupt fluctuations. The model tends to smooth these fluctuations, which can lead to less accurate predictions at points of rapid variation.

Interpretation:

The LSTM_SARIMA hybrid model demonstrates a notable ability to understand the general trends of the data. By combining the strengths of the LSTM, which captures long-term dependencies, and the SARIMA, adapted to seasonal components, this model offers an effective approach for modeling complex time series. However, it may experience difficulties in predicting rapid fluctuations and pronounced peaks, probably due to the nature of the LSTM, which is designed to capture long-term temporal dependencies but may be less responsive to short-term variations

The discrepancies between the predictions and the actual values can be explained by various factors, such as the inherent noise in the data, the complexity of the time series, or limitations specific to the LSTM model.

From a practical point of view, the accuracy of the LSTM_SARIMA model's predictions may be sufficient for some applications, but insufficient for others requiring greater accuracy.

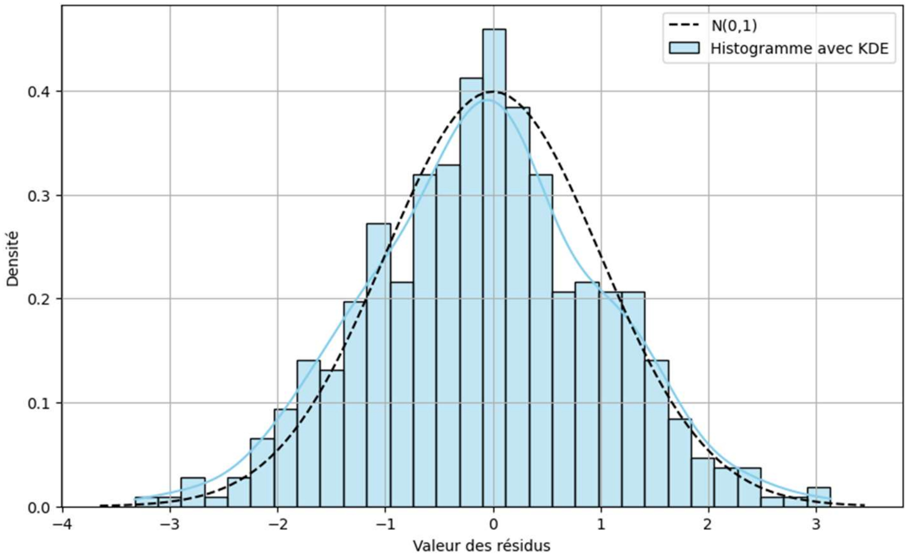


Figure 12. Histogram of Residuals with Kernel Density Estimation and N(0,1) Curve - LSTM_SARIMA

Both the simple LSTM and the LSTM_SARIMA hybrid models exhibit residuals that are generally centered around zero, adopting a bell shape that suggests a distribution close to normal. The kernel density estimation (KDE) provides a smoothed representation of the residual distribution, confirming the bell shape and slight skewness. The standard normal distribution curve (N(0,1)) serves as a reference, and it is observed that the residual distribution (KDE) generally follows it, although with some deviations. The tails of the residual distribution appear to be slightly thicker than those of N(0,1), indicating a higher probability of observing extreme values.

The LSTM_SARIMA hybrid model integrates the SARIMA model to capture the LSTM residuals, with the aim of improving the normality of these residuals. Visually, the histogram of the hybrid model's residuals seems to approach the standard normal distribution more closely compared to that of the simple LSTM model (figure 2.13), suggesting a beneficial effect of this combined approach.

In summary, while the simple LSTM model presents generally normal residuals with some deviations, the integration of the SARIMA model in the hybrid approach aims to improve this normality. Visual observations indicate that this hybridization has contributed to a slight improvement in the normality of the residuals compared to the LSTM model alone.

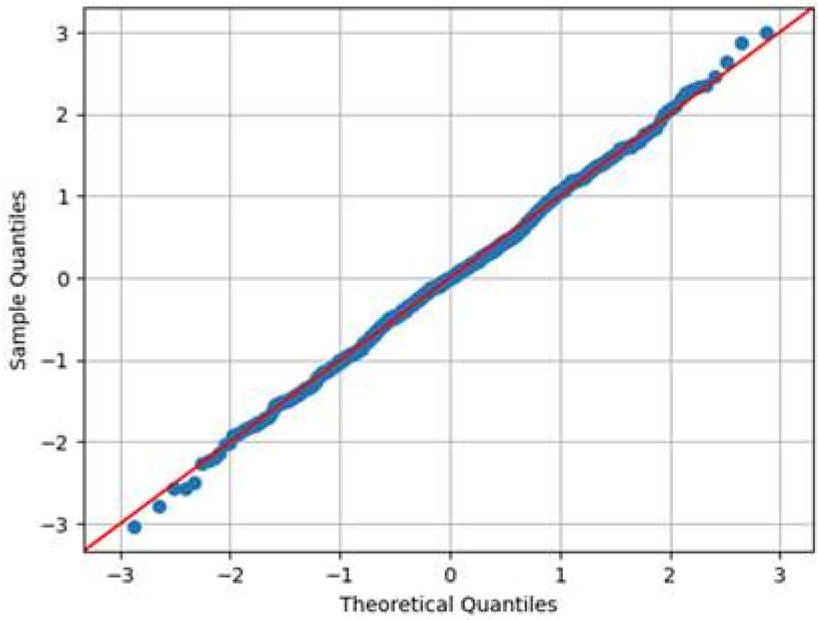


Figure 13. Q-Q Plot of Residuals – LSTM_SARIMA

The blue points, representing the quantiles of the observed residuals, are remarkably aligned with the red line, symbolizing a perfect normal distribution. This precise alignment across the entire range of values, including in the tails of the distribution, indicates that the residuals of the LSTM_SARIMA model follow an almost perfectly normal distribution. This normality of the residuals is a very positive indicator of the model's performance, reinforcing confidence in the validity of its results and forecasts. It appears that the integration of the SARIMA model to capture the LSTM residuals has succeeded in producing residuals very close to a normal distribution.

In comparison, although both models, simple LSTM and LSTM_SARIMA, exhibit residuals generally close to normal with a close alignment of the points on the red line in the Q-Q plot, the alignment is significantly more precise in the hybrid model. In the simple LSTM model, some slight deviations are visible in the tails of the distribution, while in the LSTM_SARIMA model, these deviations are almost non-existent, indicating a significant improvement in the normality of the residuals. This suggests that the addition of the SARIMA model has been very effective in improving the normality of the residuals.

3) Statistical Analysis

- Autocorrelation Analysis

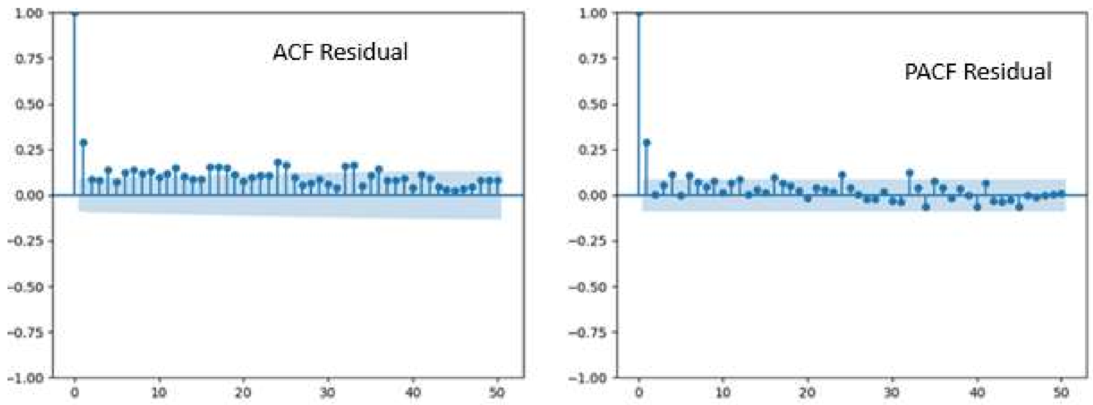


Figure 14. ACF and PACF Plots of the Residuals from the Hybrid LSTM_SARIMA Model

The graph representing the ACF of the residuals reveals that the majority of the points are within the confidence interval, except for the first lag, which is always equal to one. A few points are slightly outside this interval, but they remain very close to the limit. No clear pattern of significant autocorrelation at higher lags is observed. This absence of significant autocorrelation is a positive sign, suggesting that the LSTM_SARIMA model has successfully captured the temporal dependencies in the data, making the residuals independent. The few points slightly outside the interval could indicate a slight residual autocorrelation, but it is probably weak and does not significantly affect the model's performance.

The PACF graph of the residuals shows a similar situation. As with the ACF, most points are within the confidence interval, except for the first lag. A few points are slightly outside this interval, but they remain very close to the limit. No clear pattern of significant partial autocorrelation at higher lags is observed. The absence of significant partial autocorrelation confirms the absence of significant overall autocorrelation, reinforcing the conclusion that the LSTM_SARIMA model has well captured the temporal dependencies. The few points slightly outside the interval could indicate a slight residual partial autocorrelation, but it is probably weak and does not significantly affect the model's performance.

The analysis reveals a significant improvement of the LSTM_SARIMA model compared to the simple LSTM model, particularly with respect to the autocorrelation of the residuals. The absence of noticeable autocorrelation in the LSTM_SARIMA model demonstrates the effectiveness of combining the two models to capture temporal dependencies. While the simple LSTM model had difficulty modeling short-term dependencies, as evidenced by the significant peaks observed in the ACF and PACF graphs at lag 1, the LSTM_SARIMA model has overcome this limitation. The absence of autocorrelation in the residuals of the LSTM_SARIMA model is a positive indicator of its performance and reinforces confidence in its forecasts, while the presence of autocorrelation at lag 1 in the simple LSTM model indicates potential for improvement. These observations demonstrate the effectiveness of combining LSTM and SARIMA models to better capture temporal dependencies and improve the quality of residuals.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

- Ljung-Box :

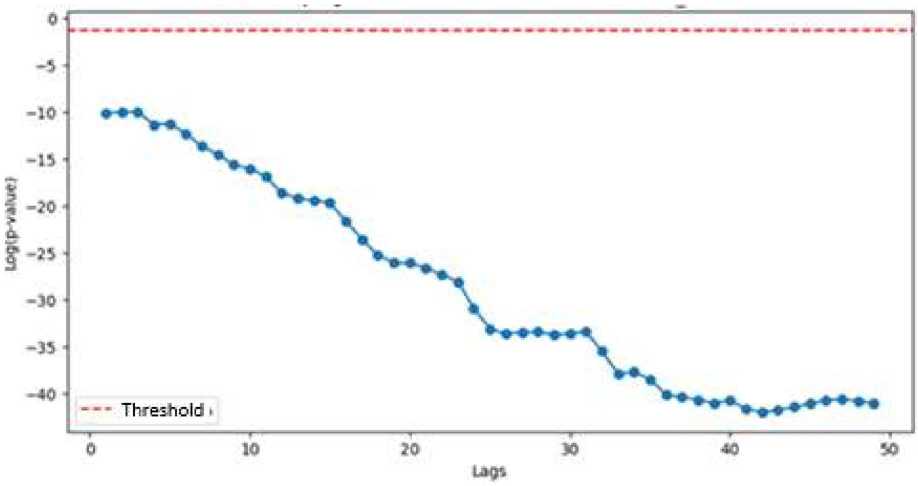


Figure 15. Log(p-value) Curves - LSTM_SARIMA

The graph in Figure 15 presents the result of the Ljung-Box test applied to the residuals of the LSTM_SARIMA model, illustrating the evolution of the logarithm of p-values as a function of lags. A dotted red line indicates the 5% significance threshold (Log (p-value) = -1.301). It is observed that the line representing the logarithms of p-values is largely below this threshold across the entire range of lags. This means that the p-values are extremely close to zero, indicating a statistically very significant autocorrelation in the residuals of the LSTM_SARIMA model.

A low autocorrelation observed in the ACF and PACF analyses may not contradict a strong autocorrelation detected by the Ljung-Box test. Indeed, ACF and PACF analyze autocorrelation at specific lags individually, while the Ljung-Box test assesses autocorrelation over a set of lags globally, thus offering a more integrated perspective. Even if individual autocorrelations do not appear significant, their combined effect may be, which the Ljung-Box test is able to detect. This underscores the importance of jointly using graphical tools and formal statistical tests for a comprehensive evaluation of autocorrelation in a time series.

- Analysis of Skewness and Kurtosis

Skewness : 0.010258068041003897

This value is extremely close to 0, which indicates an almost perfectly symmetrical distribution of the residuals. This suggests that the residual distribution of the LSTM_SARIMA model is even more symmetrical than that of the simple LSTM model.

Kurtosis : -0.002343044701141217

This value is also very close to 0, which indicates that the residual distribution has tails very similar to those of a normal distribution (rather than slightly lighter). This suggests that the residual distribution of the LSTM_SARIMA model is even closer to a normal distribution in terms of kurtosis than that of the simple LSTM model.

Comparison with the simple LSTM model:

The analysis reveals a notable improvement in the normality of the residuals in the LSTM_SARIMA model compared to the simple LSTM model. The skewness and kurtosis values, which measure the asymmetry and kurtosis of the residual distribution, respectively, are even closer to zero in the LSTM_SARIMA model. This increased proximity to zero indicates that the residual distribution of the hybrid model is closer to an ideal normal distribution than that of the simple LSTM model. Therefore, the capture of LSTM residuals by the SARIMA model seems to have effectively contributed to improving the normality of the residuals. Although both models show residuals generally close to a normal distribution, the LSTM_SARIMA model demonstrates a significant improvement in this normality. This improvement reinforces confidence in the validity of the results and forecasts of the hybrid model. In summary, the LSTM_SARIMA model displays residuals even closer to a normal distribution than the simple LSTM model, which demonstrates the effectiveness of capturing LSTM residuals by the SARIMA model to refine the normality of the residuals.

- Breusch-Pagan test for heteroscedasticity:

Test statistic : 0.3728

P-value : 0.5415

The test statistic (0.3728) is a measure of the deviation from the null hypothesis of homoscedasticity. The p-value (0.5415) indicates the probability of observing a test statistic as extreme (or more extreme) if the null hypothesis were true.

The p-value (0.5415) is significantly greater than the significance threshold of 0.05. This means that we cannot reject the null hypothesis. Thus, there is no significant statistical evidence of heteroscedasticity in the residuals of the LSTM_SARIMA model.

D. SARIMA_LSTM

1) RMSE

For the SARIMA_LSTM hybrid model, the RMSE on the test dataset is 0.9369. This means that, on average, the prediction error of this model compared to the actual temperatures is approximately 0.9369 °C, indicating an improvement over all previous models.

The analysis shows that the hybrid models, combining LSTM and SARIMA, offer superior performance compared to the individual models, whether LSTM alone or SARIMAX alone. Among these hybrid models, SARIMA_LSTM stands out as the most performant. This observation suggests that capturing the residuals of the SARIMA model by the LSTM model is the most effective strategy to improve prediction accuracy in this specific context. In contrast, the SARIMA model used alone proves to be the least performant. Although the RMSE is improved with SARIMA_LSTM, its interpretation should not be done in isolation when evaluating the model's accuracy.

2) Residual graphical analysis

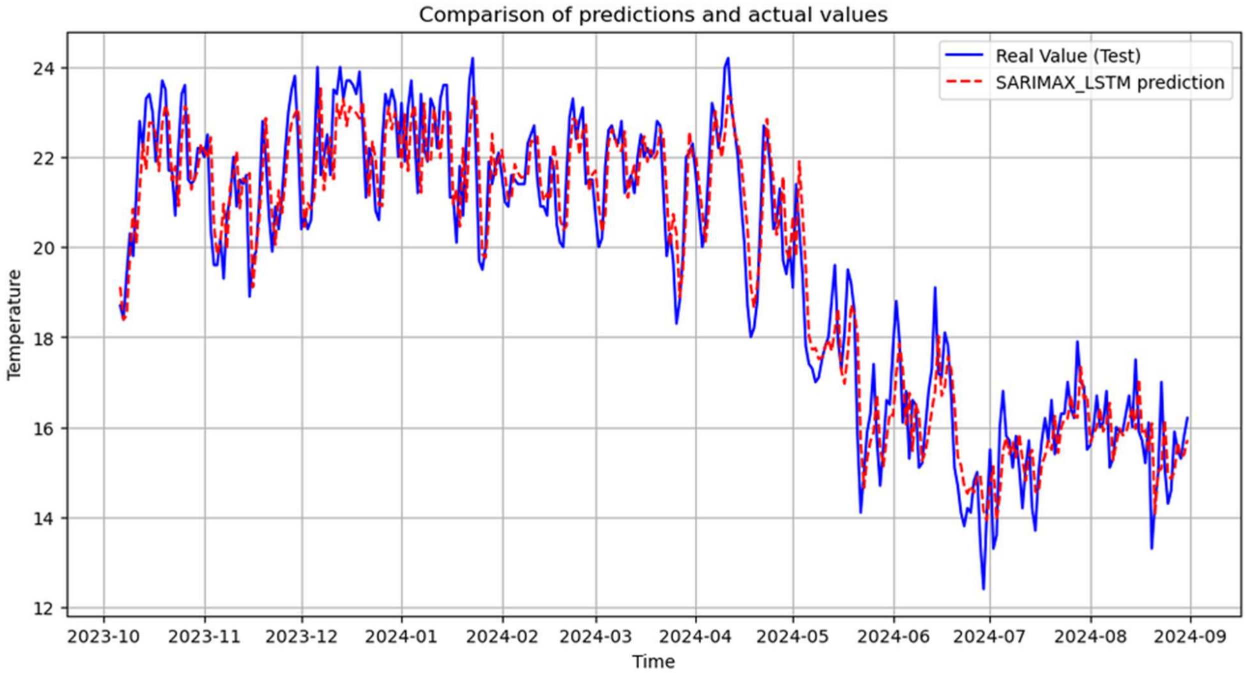


Figure 16. Difference between actual value and predicted value by the SARIMAX_LSTM hybrid model

General Observation :

The graph presents a visual comparison between the actual temperature values and the predictions of the SARIMAX_LSTM model on the test dataset. The blue curve illustrates the evolution of the actual values, revealing a time series characterized by significant fluctuations, clear trends, and a certain seasonality, with periods of temperature rises and falls, and rapid and abrupt variations. The dotted red curve represents the predictions of the SARIMAX_LSTM model, which closely follows the actual values curve, faithfully capturing the general trends and fluctuations. The striking similarity between the two curves testifies to the high performance of the SARIMAX_LSTM model, which seems to have successfully modeled the complexity of the data, including rapid variations and long-term trends, with minimal deviations between predictions and actual values.

Interpretation:

The analysis reveals that the SARIMAX_LSTM model is very effective in predicting temperatures on the test dataset, efficiently capturing the trends, fluctuations, and seasonality of the data. The combination of SARIMA and LSTM seems particularly effective in modeling this time series. The accuracy of the predictions, with few deviations from the actual values, suggests that the model is well-fitted to the data and capable of providing reliable forecasts. The model has demonstrated a notable ability to grasp the complex dependencies present in the data, encompassing both rapid variations and long-term trends. This performance suggests that integrating the SARIMA model with the LSTM network to predict residuals is an effective approach for modeling time series with varied characteristics.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

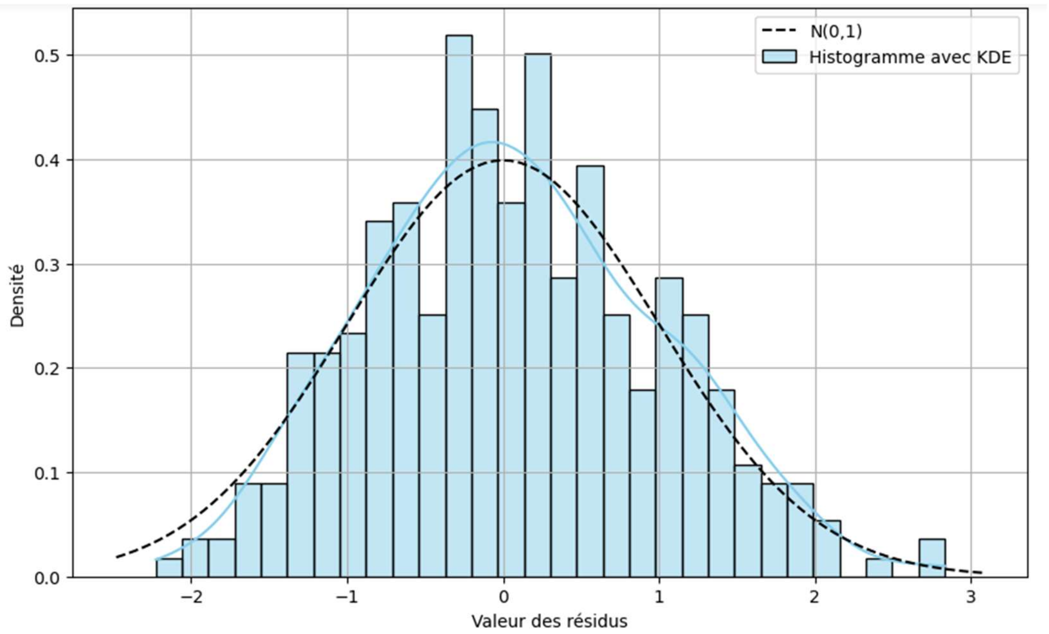


Figure 17. Histogram of residuals with kernel density estimates and N(0,1) curve - SARIMA_LSTM

General observation :

The graph reveals that the residuals of the SARIMA_LSTM hybrid model exhibit a distribution that approximates normal, as evidenced by the bell shape of the histogram and the proximity of the kernel density estimation (KDE) curve to the standard normal distribution curve N(0,1). This suggests that the model has effectively captured the patterns present in the data, leaving behind errors that behave randomly and Gaussianly.

Interpretation :

Comparing with the LSTM_SARIMA hybrid model, we observe a striking similarity in the residual distribution. Both hybrid approaches seem to have succeeded in producing residuals that closely align with a normal distribution. This indicates that, regardless of the order of application of the LSTM and SARIMA models, the combination of the two results in residuals that behave satisfactorily.

However, it is important to note that the normality of the residuals alone does not guarantee the perfection of the model. Other factors must also be taken into account. Nevertheless, the normal distribution of the residuals is a positive indicator of the quality of the model fit and reinforces confidence in its ability to provide reliable forecasts.

International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

https://ijpsat.org/
Vol. 50 No. 1 April 2025, pp. 69-96

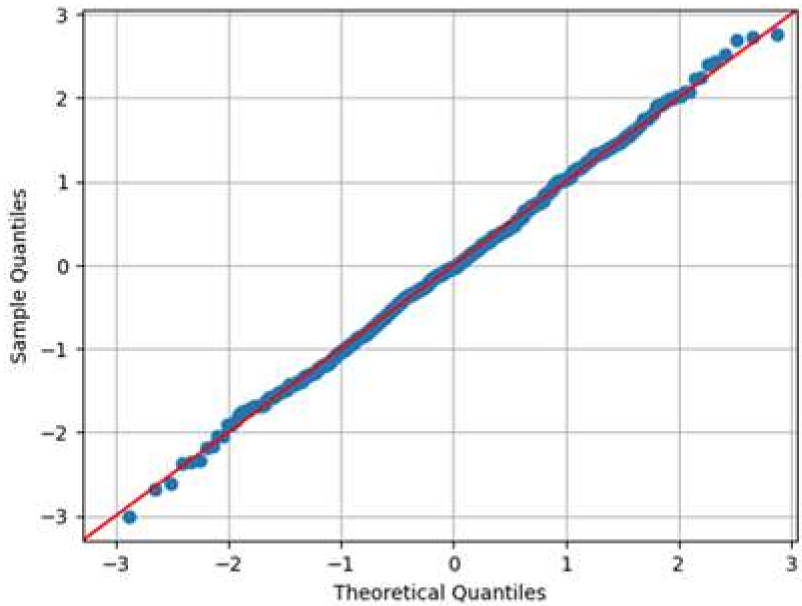


Figure 18. Q-Q Plot of Residuals - SARIMA_LSTM

General Observation :

This Q-Q plot, which compares the quantiles of the observed residuals with those of a theoretical normal distribution, reveals a strong agreement. The points representing the residuals closely align along the reference line, indicating that the residuals of the SARIMA_LSTM model behave very similarly to a normal distribution.

This observation is crucial because it suggests that the model has effectively captured the patterns present in the data, leaving behind errors that behave randomly and Gaussianly.

Interpretation :

The SARIMA_LSTM hybrid approach succeeds in producing residuals that conform satisfactorily to a normal distribution.

Just like the LSTM_SARIMA model, the SARIMA_LSTM model demonstrates a comparable ability to model time series in such a way that the residual errors behave as expected with a normal distribution. This absence of significant difference in the residual distribution between the two models suggests that the order of application of the LSTM and SARIMA models does not have a major impact on the normality of the residuals.

3) Statistical analysis

- Autocorrelation analysis

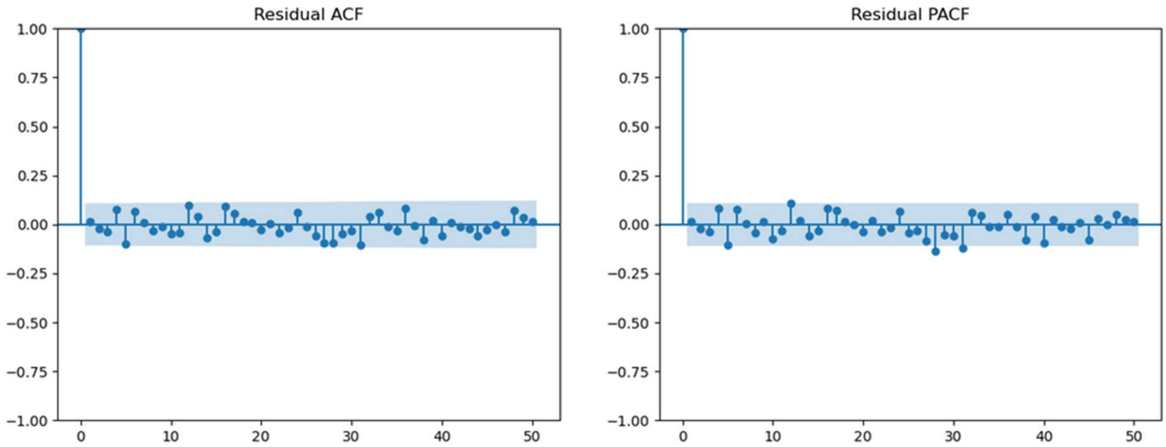


Figure 19. ACF and PACF Plots of Residuals from the Hybrid SARIMAP_LSTM Model

The examination of the autocorrelation (ACF) and partial autocorrelation (PACF) functions of the residuals of the SARIMA_LSTM model reveals a notable absence of significant structure. The vast majority of points, in both the ACF and PACF, are located within the shaded area, indicating that the autocorrelations and partial autocorrelations of the residuals are not statistically significant.

In other words, the residuals behave randomly and do not exhibit significant temporal or partial dependencies. This absence of structure is a positive indicator of the quality of the model fit, as it suggests that the model has well captured the temporal dependencies in the data, leaving behind errors that behave like white noise.

Therefore, the ACF and PACF graphs reinforce confidence in the validity and reliability of the model's forecasts, as they confirm that the residuals are random and do not contain additional information that could be modeled.

Thus, both hybrid models, SARIMA_LSTM and LSTM_SARIMA, generate residuals that behave satisfactorily. In both cases, the majority of points are located within the confidence interval, indicating a general absence of significant autocorrelation. However, a subtle difference emerges in the treatment of the first lag. The SARIMA_LSTM model does not exhibit significant autocorrelation at this lag, while the LSTM_SARIMA model does. This observation suggests that the SARIMA_LSTM model might be more effective in capturing the short-term dependencies present at lag 1.

- Ljung-Box :

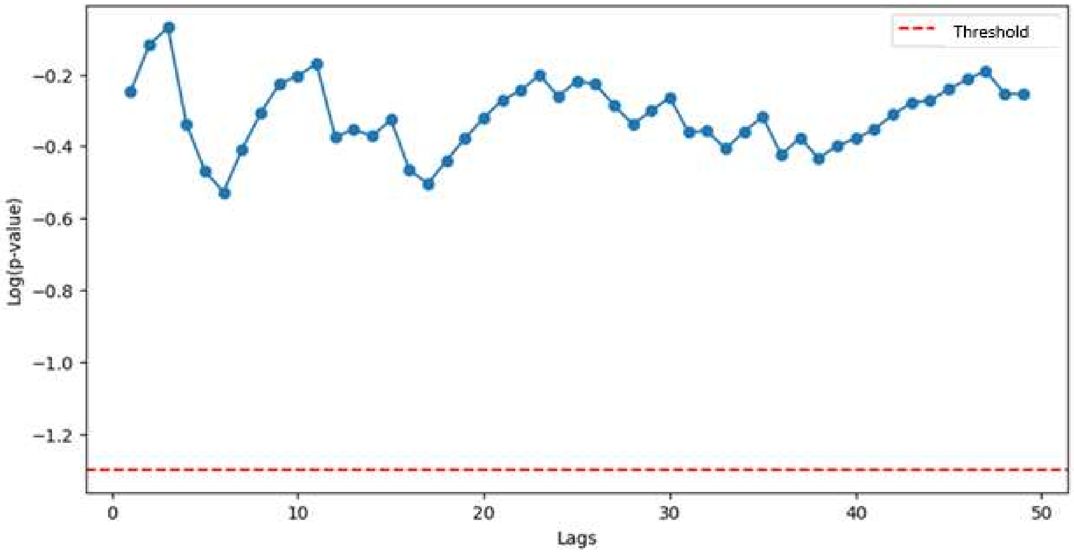


Figure 20. Log(p-value) Curves – SARIMA_LSTM

The analysis of the Ljung-Box test, represented by this graph comparing the p-values of the residuals at time lags with a 5% significance threshold, reveals an absence of significant autocorrelation in the residuals of the SARIMA_LSTM model. All p-values, expressed in logarithm, are above the 5% threshold, which means they are greater than 0.05.

In other words, the Ljung-Box test fails to reject the null hypothesis of independence of the residuals. This absence of significant autocorrelation is a positive indicator of the quality of the model fit, as it suggests that the model has well captured the temporal dependencies in the data, leaving behind errors that behave randomly.

- Skewness and kurtosis analysis

Skewness : 0.20481312412300262

In this case, the skewness is 0.205, which is slightly greater than 0. This indicates a slight positive skewness in the distribution of the residuals, with a slightly longer tail on the right. Although this value is slightly farther from 0 than in the example you provided, it remains relatively close to 0, which suggests that the distribution of the residuals is generally symmetrical. This slight skewness is often considered acceptable and does not necessarily indicate a major problem with the model.

Kurtosis : -0.2751044018408244

The kurtosis is -0.275, which is less than 3. This means that the distribution of the residuals has slightly lighter tails than those of a normal distribution (platykurtosis). This slight platykurtosis is generally considered acceptable and does not indicate a major problem with the model.

- Breusch-Pagan Test for Heteroscedasticity:

Statistic test: 0.1185

P-value : 0.7307

The obtained p-value (0.7307) is greater than the commonly used significance level (0.05). This means that we cannot reject the null hypothesis. In other words, there is not enough statistical evidence to conclude that the residuals of the SARIMA_LSTM model exhibit significant heteroscedasticity. The absence of heteroscedasticity in the SARIMA_LSTM model suggests that the variance of the residuals is constant and that there is no significant relationship between the explanatory variables and the dispersion of the errors.

V. SYNTHESIS

The evaluation of the simple LSTM and SARIMA models highlighted their limitations in predicting the temperature data of our study. Although the LSTM model demonstrated superior performance compared to the SARIMA model, it is important to note that traditional time series models like SARIMA can struggle to capture the complex non-linearities and long-term dependencies present in temperature data. The LSTM, on the other hand, is designed to handle these complexities thanks to its ability to learn complex sequential patterns.

The results of the residual analysis, both graphical and statistical, confirmed the superiority of the LSTM model over the SARIMA model. The LSTM model's residuals showed a distribution closer to normal and lower autocorrelation, indicating a better ability to capture the relevant information in the data.

However, given the inherent weaknesses of the simple LSTM model, we explored hybrid approaches combining LSTM and SARIMA. The LSTM_SARIMA and SARIMA_LSTM hybrid models showed similarities in the initial analyses, particularly regarding the residual distribution (KDE curve) and the normality of the residuals (QQ curve).

Nevertheless, the deviation curve between the predicted and actual values for the SARIMA_LSTM model proved particularly impressive, revealing an almost perfect similarity between the two curves. This suggests that the SARIMA_LSTM model successfully captured the patterns present in the data with high accuracy, making it a promising model for predicting temperature data.

The more in-depth comparative evaluation of the LSTM_SARIMA and SARIMA_LSTM hybrid models highlighted significant differences in their ability to generate residuals that behave like white noise. The SARIMA_LSTM model stood out by successfully validating both autocorrelation tests, namely the autocorrelation (ACF) and partial autocorrelation (PACF) functions, as well as the Ljung-Box test.

It should be specified that these tests evaluate autocorrelation from different perspectives. The ACF and PACF functions analyze autocorrelation at specific lags, allowing the identification of residual temporal dependencies at particular lags. In contrast, the Ljung-Box test is a global approach aimed at detecting the presence of significant autocorrelation over a set of lags, thus evaluating the hypothesis of residual independence over the considered time horizon.

In the case of the LSTM_SARIMA model, the results of the Ljung-Box test revealed significant residual autocorrelation, suggesting an inability of the model to capture all the temporal structures present in the data. This persistence of unmodeled temporal dependence indicates the existence of predictable patterns not taken into account by the LSTM_SARIMA approach.

Conversely, the SARIMA_LSTM model produced residuals free of significant autocorrelation, both at the individual lag level (ACF and PACF) and at the global level (Ljung-Box). These results suggest that the SARIMA_LSTM model better understood the underlying temporal dynamics, resulting in residuals assimilable to white noise and thus confirming a better ability to model temporal dependencies.

However, the comparative analysis of the LSTM_SARIMA and SARIMA_LSTM hybrid models reveals a notable trade-off between the normality of the residuals and the absence of autocorrelation. The LSTM_SARIMA model demonstrates a remarkable fit to an ideal normal distribution, illustrated by an almost perfect symmetry and a kurtosis similar to that of a normal law. However, it shows signs of residual autocorrelation, detected by the Ljung-Box test, suggesting that temporal dependencies have not been fully captured.

Conversely, the SARIMA_LSTM model excels in eliminating autocorrelation, as evidenced by the PACF, ACF, and Ljung-Box tests. However, it exhibits a slight deviation from ideal normality, with a slight positive skewness and platykurtosis of the residuals.

In our context, the absence of autocorrelation is a determining criterion for ensuring the robustness of long-term forecasts and ensuring a faithful modeling of the temporal dynamics of the series. The slight positive skewness and platykurtosis observed remain within acceptable limits, which leads to favoring the SARIMA_LSTM hybrid model.

We must still emphasize that both models demonstrate homoscedasticity of the residuals, which reinforces their reliability for forecasts.

VI. FUTURE DIRECTIONS

The perspectives of this study are oriented towards several areas of improvement. Integrating new data sources, whether climatic, geographic, or socio-economic, could refine the accuracy of forecasts. By taking into account parameters such as humidity, atmospheric pressure, wind speed, or even urbanization, the model could better capture the dynamics influencing temperature variations in Antananarivo.

Exploring more advanced models also represents a promising avenue. Using convolutional neural networks would allow extracting complex spatio-temporal patterns, while attention mechanisms could help identify the most influential variables.

Applying the model to other regions and at different time scales represents another challenge. A comparative study conducted in varied climatic contexts would assess the generalizability of the developed approaches. Furthermore, adapting the model for forecasts ranging from very short-term to long-term would open up broader usage perspectives.

Finally, developing practical tools would facilitate the dissemination of forecasts to the public and decision-makers. Implementing a web or mobile application, as well as integrating the model into early warning systems for extreme weather events, would promote better anticipation of climate risks and more informed decision-making.

VII. CONCLUSIONS

In this research, we were able to compare meteorological forecasting techniques integrating various methodologies ranging from statistical models to advanced machine learning algorithms. Traditional approaches, such as the SARIMA model, leverage historical patterns to predict future weather conditions by analyzing time series with seasonal components. Conversely, modern techniques, like Long Short-Term Memory Recurrent Neural Networks (LSTM RNN), take advantage of deep learning to capture complex and non-linear relationships in the data, thereby enhancing predictive capabilities in the face of variable climatic conditions.

The results obtained with meteorological data from Antananarivo allowed us to compare and evaluate various modeling approaches for temperature forecasting, highlighting the strengths and weaknesses of individual LSTM and SARIMA models, as well as their hybrid combinations.

Our results demonstrated that hybrid models, particularly SARIMA_LSTM, outperform simple models in terms of accuracy and reliability. The in-depth evaluation of residuals revealed that the SARIMA_LSTM model, integrating exogenous variables into the SARIMA model and using LSTM to refine forecasts, generated residuals behaving like white noise, thus validating its ability to capture the complex temporal dynamics of temperature data.

Although the LSTM_SARIMA model demonstrated a better fit to an ideal normal distribution, it exhibited significant residual autocorrelation, which limits its ability to provide robust long-term forecasts. In contrast, the SARIMA_LSTM model, despite a slight deviation from normality, excelled in eliminating autocorrelation, an essential criterion for ensuring forecast reliability.

In the specific context of our study, where the robustness of long-term forecasts is paramount, the SARIMA_LSTM hybrid model proved to be the most performant. Its ability to faithfully model the temporal dynamics of the data, while maintaining residuals free of significant autocorrelation, makes it a valuable tool for improving the accuracy of meteorological forecasts in Antananarivo.

Finally, it is important to emphasize that both hybrid models demonstrated homoscedasticity of the residuals, thus reinforcing their reliability for forecasts. This study contributes to the advancement of temperature forecasting techniques by proposing an effective hybrid approach, adapted to the climatic peculiarities of Antananarivo but also for other regions.

REFERENCES

[1] Box,G.E.P.,Jenkins,G.M.,Reinsel,G.C.,&Ljung,G.M. (2015). Time Series Analysis: Forecasting and Control (5th Edition). Wiley.

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

[3] Brockwell, P. J., & Davis, R. A., 1991, Time Series: Theory and Methods (2nd ed.), Springer-Verlag.

[4] Arthus Anin, 2024, "Séries temporelles : SARIMA pour initiation aux prédictions temporelles", Ivoire Talents Labs

[5] Veysel Süleyman Yavuz, 7 feb. 2025, " Forecasting monthly rainfall and temperature patterns in Van Province, Türkiye, using ARIMA and SARIMA models: a long-term climate analysis", Journal of Water and Climate Change, Volume 16, Numéro 2, pages 800–818

[6] Sepp Hochreiter, Jürgen Schmidhuber, 1997, "Long Short-Term Memory," Neural Computation, Volume 9, Issue 8, pages 1735–1780.

[7] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). «A Critical Review of Recurrent Neural Networks for Sequence Learning », arXiv preprint, arXiv : 1506.00019.

[8] Sepp Hochreiter, Jürgen Schmidhuber, 1997, "Long Short-Term Memory," Neural Computation, Volume 9, Issue 8, pages 1735–1780.

[9] Ljung, G. M., & Box, G. E. P. (1978). On a measure of a lack of fit in time series models. Biometrika, 65(2), 297-303. https://doi.org/10.1093/biomet/65.2.297

[10] Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting. Springer Science & Business Media, pp. 292-298.

[11] Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. MIT Press, pp. 267-275.

[12] Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th ed.). Pearson Education, pp. 67-72.