

Article

Not peer-reviewed version

Wavelet-Enhanced Deep Learning for Multi-Variable Meteorological Time-Series Forecasting in Togo

[Apeke Sena](#)*, Gbafa Senanou, Wotodzo Kokou Felicien, [Kpogo-Nuwoklo K. Agbéko](#), [Agboka Komi](#), [Ouro-Djobo Sanoussi S.](#)

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0851.v1

Keywords: deep learning; model; SWT; MHA; LSTM; temperature; relative humidity; wind speed



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

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

Wavelet-Enhanced Deep Learning for Multi-Variable Meteorological Time-Series Forecasting in Togo

Apeke Sena ^{1,2,*}, Gbafa Senanou ^{2,3}, Wotodzo Kokou Felicien ³, Kpogo-Nuwoklo K. Agbéko ⁴, Agboka Komi ⁵ and Ouro-Djobo Sanoussi S. ^{2,6}

¹ Laboratoire de Recherche en Science de l'Ingénieur (LARSIS); Université de Lomé, EPL-UL, Lomé1, 01BP 1515, Togo

² Centre d'Excellence Régional pour la Maîtrise de l'Electricité (CERME), University of Lome, Lome, P.O. Box 1515, Lome Togo

³ Laboratoire de Structures et de Mécanique des Matériaux (LaS2M) ; Université de Lomé, EPL-UL, Lomé1, 01BP 1515, Togo

⁴ Météo-France, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France

⁵ West African Science Center on Climate and Adapted Land use (WASCAL), Togo

⁶ Solar Energy Laboratory, Department of Physics, Faculty of Science, University of Lomé, Lomé, Togo

* Correspondence: apekekodjo@gmail.com; Tel.: +22893371331

Abstract

Accurate short-term forecasting of key meteorological variables—air temperature, relative humidity, and wind speed—remains challenging in tropical and sub-Saharan regions due to strong diurnal cycles, seasonal variability, and non-stationary dynamics. To address these limitations, this study proposes a hybrid deep learning model combining Stationary Wavelet Transform (SWT), Multi-Head Attention (MHA), and LSTM networks. First, SWT decomposes meteorological time series into multi-scale components, capturing both low-frequency trends and high-frequency fluctuations while preserving temporal resolution. Then, the attention mechanism dynamically weights the importance of these multi-scale features across time, enhancing the model's ability to focus on the most relevant patterns and interactions. Finally, LSTM layers model long-term dependencies and nonlinear temporal structures to generate multi-output predictions. The model is trained on hourly data enriched with lagged and statistical features. Experimental results show strong predictive performance (MAE = 1.21, RMSE = 2.01, $R^2 = 94\%$), with notable improvements in modeling rapid variations, especially for wind speed and humidity. This work represents one of the first integrations of SWT, attention mechanisms, and LSTM for multi-variable forecasting in tropical climates, with practical applications in energy forecasting and precision agriculture.

Keywords: deep learning; model; SWT; MHA; LSTM; temperature; relative humidity; wind speed

1. Introduction

Diurnal and seasonal variations in meteorological parameters such as air temperature, relative humidity, and wind speed are mainly the result of the Earth's rotation on its axis around the Sun. In general, the temperature rises after sunrise, reaching a maximum in the middle or late afternoon, while relative humidity often shows the opposite trend (dilution of moisture for an almost constant vapor content). Wind speed is also influenced by atmospheric stability and surface-boundary layer coupling, leading to contrasting diurnal regimes depending on synoptic conditions and surface roughness [1–3]. In West Africa, this variability is structured by seasonal monsoon dynamics, the migration of the Intertropical Convergence Zone (ITCZ), and regional factors (advected moisture, cloud cover, land-sea gradients, and vegetation variability). In Togo, the south (Guinean coastal influence) and the north (Sudano-Sahelian influence) can have distinct seasonal cycles, with dry and dusty episodes associated with the harmattan and wet periods where relative humidity and cloud cover dominate. These spatial and seasonal contrasts affect evapotranspiration, perceived temperature, and wind conditions, and are often amplified by interannual variability and extreme events (heat waves, rainfall deficits, convective gusts) [4–7]. This variability has direct socioeconomic consequences for many sectors. In agriculture, it

determines crop calendars, water and heat stress, pest and disease dynamics, and therefore the stability of yields and incomes. In the energy sector, it influences both demand (air conditioning, cooling) and renewable supply (wind, solar), while in the building and transport sectors, it affects the planning, operation, and resilience of infrastructure. For human health, exposure to heat and humidity (heat index) can increase morbidity and mortality, and climate variability interacts with environmental and socio-economic factors to modulate health risks [8–10].

Faced with these challenges, modeling and forecasting have become strategic tools for anticipating, mitigating, and managing risks. Historically, the dominant approaches have been based on deterministic physical models, founded on the equations of fluid dynamics and thermodynamics (conservation of mass, momentum, and energy), solved numerically on spatio-temporal grids. Numerical prediction systems and climate models (global or regional) form the backbone of meteorological services, but their local performance depends heavily on the quality of the initial conditions, parameterization schemes (convection, clouds, turbulence), resolution, and representativeness of assimilated observations. Furthermore, the chaotic nature of the atmosphere imposes limits on predictability, which justifies the quantification of uncertainty [11,12].

To complement these models, numerous statistical methods have been used to produce rapid and interpretable forecasts adapted to time series, particularly for short horizons or sectoral applications. AR, MA, ARMA, and ARIMA models, as well as their seasonal variants (SARIMA), are benchmarks for modeling temporal dependencies and seasonal structures. Other families include state space models, exponential smoothing approaches, and regression techniques (linear, nonlinear) with exogenous variables. However, these methods often assume a form of stationarity and linearity, and may be erroneous in the presence of nonlinearities, regime shifts, extremes, or complex spatio-temporal dependencies [13–16].

The rise of machine learning has subsequently enabled the incorporation of nonlinear relationships and complex interactions between predictors, while leveraging growing volumes of data from stations, reanalyses, and remote sensing. Methods such as random forests, support vector machines, and reinforcement learning algorithms have performed well in various meteorological regression and environmental variable prediction tasks, particularly when explicit physical relationships are difficult to parameterize at high resolution. However, these approaches remain sensitive to data quality (noise, missing data), distribution bias (domain shift), variable selection, and may suffer from limited interpretability or difficulty in simultaneously capturing spatial structures and long-range temporal dependencies [17–20].

More recently, deep learning has provided specialized architectures for exploiting the structure of spatio-temporal data. Convolutional Neural Networks (CNNs) are suitable for extracting spatial patterns on grids (temperature maps, humidity, wind fields), while recurrent networks (Long Short-Term Memory (LSTM)/Gated Recurrent Unit (GRU)) effectively model temporal dependencies, including long-term ones. Hybrid spatio-temporal models (e.g., ConvLSTM) unify these two dimensions by processing card sequences, and Transformers have extended these capabilities via multi-head attention mechanisms (MHA) capable of capturing long-range interactions and non-local dependencies. Despite notable progress, these models require large volumes of data, can be expensive to train, and pose challenges of explainability and physical consistency (compliance with conservation constraints, plausibility of extremes) [21–23].

Weather series (temperature, humidity, wind, rain) in the West African tropical zone are non-stationary with multi-scale seasonality, extreme events (storms, harmattan, monsoon), and sometimes noisy and/or missing data (rare/irregular stations). [24,25] have done interesting work on tropical data. In this work, we wish to take local variations into account with the main objective of designing a hybrid approach explicitly combining: (i) a “Wavelet Transform” module capturing local spatio-temporal characteristics, (ii) a module using MultiHead Attention to model global and adaptive spatio-temporal dependencies (dynamic weighting of relevant regions and instants), and (iii) a final sequential “LSTM” module focused on future dynamics and the generation of predictive trajectories.

This architecture aims to leverage the strengths of time-frequency and attentional representations, while maintaining robust sequential forecasting capabilities. This hybrid model has been used to forecast the following meteorological parameters: temperature, humidity, and wind speed with lead times of up to 72 hours and 120 hours. This approach is particularly relevant for local applications in Togo and West Africa, where spatial heterogeneity, seasonal patterns, and variable availability of observations require models that are expressive, robust, and parsimonious.

2. Materials and Methods

2.1. Data Collection and Preprocessing

Data from synoptic stations and rain gauge stations in Togo's meteorological network (see Figure 1 showing these sites on a map of Togo) could not be obtained for this study. Thus, in this work, the data used are reanalysis data from the ERA5 model. ERA5 is the fifth generation of ECMWF reanalysis for climate and global meteorology of the last eight decades. Data is available from 1940 [26]. Reanalysis combines model data with observations of the world entire data set to form a globally complete and coherent data set, applying the laws of physics. Meteorological data from six cities in Togo including Lomé, Sokodé, Atakpamé, Dapaong, Kara, Kpalimé were collected over a period from 2000 to 2024 with an hourly frequency, i.e. 24 observations per day and per city in a csv format. Also, a city column has been added to the data. They contain a total of 1,315 008 lines and 38 columns (variable). An exploration and cleaning was carried out on the data. Then, feature engineering work was carried out, here are some treatments: extracting the time, day of the week, month and year from the time column. Creation of binary indicators for weekend and daytime hours. Application of sine and cosine transformations to hour variables, month and dayofweek to capture their cyclical nature. Creation of interaction variables such as the difference between temperature and dew point and the difference between apparent temperature and temperature real; Conversion of wind speed and direction at 10m into vector components; Calculation of the ratio between direct and diffuse radiation. Recursive Feature Elimination (RFE) was used with the XGBoost model to select variables that have an impact on the targets. Variables retained for multivariate prediction of air temperature, relative humidity, and wind speed are referenced in the Tables A2 and A3 (See the Appendix A part).

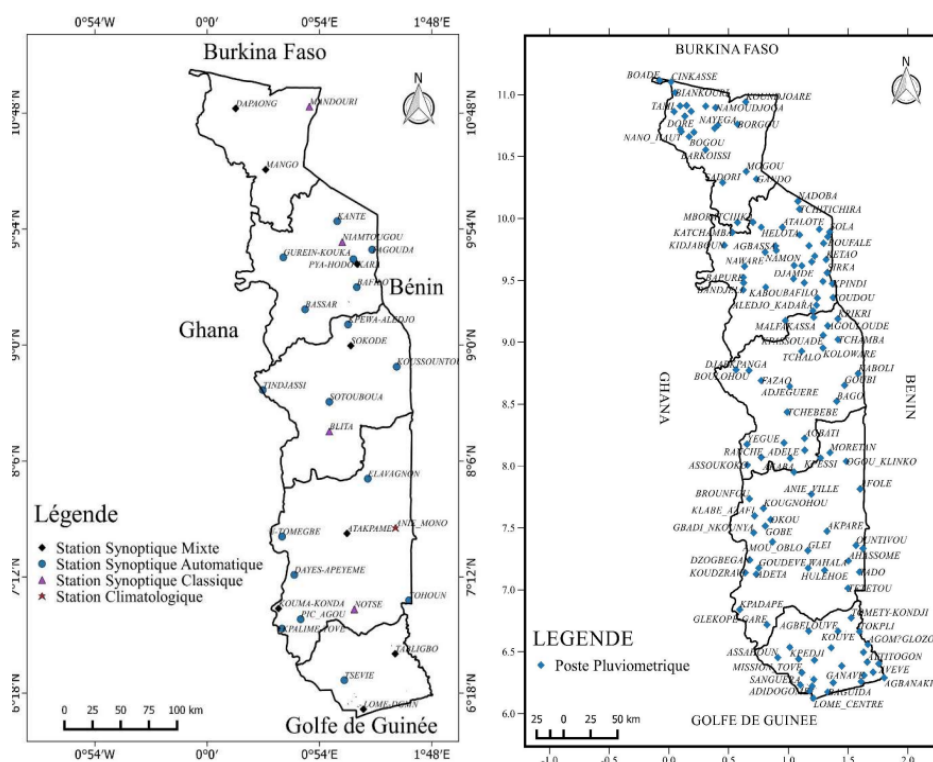


Figure 1. Synoptic stations and rain gauge stations in Togo's meteorological network [27].

2.2. Mathematical Design and Writing of the Model

Let $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ denote the set of meteorological stations, where $N = 6$ corresponds to Lomé, Kpalimé, Atakpamé, Sokodé, Kara and Dapaong (Six cities of Togo). For each city v_i , a multivariate meteorological time series is observed:

$$\mathbf{X}(t) = [x_1(t), x_2(t), \dots, x_F(t)], \quad (1)$$

where F represents the number of meteorological and temporal covariates, $X(t)$ denote the multivariate hourly observations at time step t (e.g., temperature, relative humidity, wind speed). X is the set of variables (the columns of the data set) under study. It is the same for all cities selected for this study in Togo.

Given a historical window of length L , the forecasting task aims to predict future values of air temperature at 2 m, relative humidity, and wind speed at 10 m over a forecasting horizon $H \in \{72, 120\}$:

$$\hat{\mathbf{Y}}_i(t+1:t+H) = f_\theta(\mathbf{X}(t-L+1:t)), \quad (2)$$

where f_θ denotes a nonlinear predictive function parameterized by θ . In this work, the function f_θ represents the composite of three models, the Stationary Wavelet Transform (SWT), a Multi-Head Attention (MHA) mechanism and the Long Short-Term Memory (LSTM):

$$f_\theta \equiv \text{SWT} \circ \text{MHA} \circ \text{LSTM} \quad (3)$$

The functions SWT, MHA, and LSTM in Equation (3) are described in the following subsections.

2.3. Wavelet-Based Multi-Resolution Analysis

Meteorological time series are inherently non-stationary, characterized by multiple temporal scales such as diurnal cycles, synoptic variations, and seasonal trends. To explicitly capture these dynamics, a wavelet-based multi-resolution decomposition is adopted. Given a signal $x(t)$, the Stationary Wavelet Transform (SWT) decomposes it into approximation and detail components:

$$x(t) = A_J(t) + \sum_{j=1}^J D_j(t), \quad (4)$$

where $A_J(t)$ represents the low-frequency trend and $D_j(t)$ captures fluctuations at scale j . Unlike the Discrete Wavelet Transform (DWT), the SWT preserves translation invariance, which is crucial for meteorological signals exhibiting phase shifts [28]. Daubechies wavelets (db4) are selected due to their compact support and demonstrated effectiveness in atmospheric and energy time series analysis [29].

2.4. Multi-Head Attention Mechanism

While wavelet decomposition provides a rich multi-scale representation, not all temporal scales or variables contribute equally at every time step. To dynamically select relevant features, a Multi-Head Attention (MHA) mechanism is employed.

Given an input sequence $\mathbf{I}_s \in \mathbb{R}^{L \times d}$, scaled dot-product attention is defined as in [23]:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \quad (5)$$

where queries (Q), keys (K), and values (V) are linear projections of \mathbf{H} :

$$\mathbf{Q} = \mathbf{I}_s W^Q, \quad \mathbf{K} = \mathbf{I}_s W^K, \quad \mathbf{V} = \mathbf{I}_s W^V. \quad (6)$$

Multi-head attention extends this mechanism by learning multiple attention subspaces:

$$\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_M)W^O. \quad (7)$$

where M is the number of attention heads.

This allows the model to jointly attend to short-term fluctuations, long-term dependencies, and inter-variable interactions, which is particularly relevant in meteorological forecasting [30,31]. W^l , $l \in \{(Q), (K), (V), \text{output}(O)\}$, represent the weight matrices, they are learned during model training.

2.5. Temporal Modeling Using LSTM Networks

To capture temporal dependencies and sequential continuity, the attention-refined representations are processed using Long Short-Term Memory (LSTM) networks [21]. LSTM networks are particularly effective for modeling noisy and long-term meteorological sequences, complementing the global dependency modeling capability of attention mechanisms.

2.6. Hybrid Wavelet–Attention–LSTM Architecture

The proposed hybrid architecture integrates three complementary components:

1. SWT-based wavelet decomposition for explicit multi-scale feature extraction,
2. Multi-head attention for adaptive scale and feature weighting,
3. LSTM layers for temporal sequence modeling.

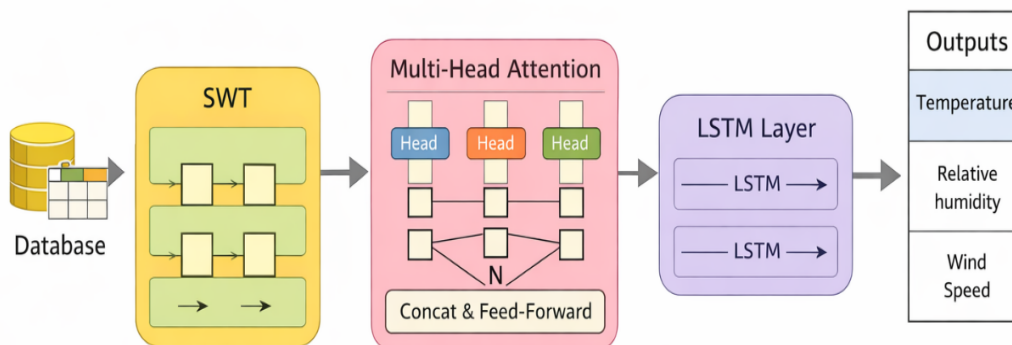


Figure 2. Architecture of the proposed hybrid model. "OpenAI" February 2026, URL: <https://chatgpt.com/g/g-2fkFE8rbu-dall-e/c/699441ee-e348-8325-b43b-2eff72c4a227>.

This design leverages the theoretical strengths of each method, as demonstrated in prior hybrid forecasting studies [32?].

The input sequence is defined as:

$$X_w = [X(t - L + 1), \dots, X(t)] \in \mathbb{R}^{L \times F}. \quad (8)$$

Each feature of X_w is independently decomposed using a Stationary Wavelet Transform (SWT) with n_d levels. For each variable, the decomposition produces one approximation component $A_{n_d}(t)$ and n_d detail components $\{D_1(t), \dots, D_{n_d}(t)\}$:

$$X(t) \rightarrow \{A_L(t), D_1(t), \dots, D_{n_d}(t)\}. \quad (9)$$

The resulting multi-scale representation is obtained by concatenating the raw input with all wavelet components:

$$Z_w(t) = \text{concat}(X(t), A_{n_d}(t), D_1(t), \dots, D_{n_d}(t)), \quad Z_w \in \mathbb{R}^{L \times F'} \quad (10)$$

where $F' = F \times (n_d + 1)$ denotes the expanded feature dimension.

The multi-scale sequence Z_w is then processed by a Multi-Head Self-Attention (MHA) mechanism. Linear projections are first applied to obtain the query, key, and value matrices:

$$Q = Z_w W_Q, \quad K = Z_w W_K, \quad V = Z_w W_V. \quad (11)$$

The outputs of multiple heads are concatenated and linearly transformed to produce the contextualized representation:

$$H = \text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_M) W^O. \quad (12)$$

H is subsequently fed into an LSTM network to capture sequential temporal dependencies:

$$h_t, c_t = \text{LSTM}(H_t, h_{t-1}, c_{t-1}), \quad (13)$$

where h_t and c_t denote the hidden and cell states, respectively.

The final hidden state h_T summarizes the temporal dynamics of the input window: $h_T \in \mathbb{R}^{d_h}$.

The final prediction for a forecasting horizon h is obtained through a dense regression layer:

$$\hat{y}(t+h) = W h_T + b, \quad (14)$$

where $\hat{y}(t+h) \in \mathbb{R}^F$ represents the multi-output forecast.

2.7. Training Objective and Evaluation Metrics

The model is trained by minimizing the Mean Absolute Error (MAE):

$$\mathcal{L}_{\text{MAE}} = \frac{1}{H} \sum_{h=1}^H |y(t+h) - \hat{y}(t+h)| \quad (15)$$

Performance is evaluated using MAE, Root Mean Square Error (RMSE), and the coefficient of determination (R^2), following standard practices in meteorological forecasting [15].

3. Results

This part presents an in-depth analysis of the performance of the proposed hybrid model, combining a wavelet decomposition, a Multi-Head Attention (MHA) mechanism and a Long Short Term Memory (LSTM). The main objective is to evaluate the model's ability to simultaneously predict several key meteorological variables, namely temperature at 2m, relative humidity at 2m and wind speed at 10m, from hourly time series enriched by derived variables (lags, rolling averages, cyclical components). The data was divided into three parts: the first part (January 2000 to December 2023) for the training, the second part (January 2024 to June 2024) for the validation, and the last part (July 2024 to December 2024) for testing. Performance is evaluated on an independent test set, not used during training or hyperparameter selection. The metrics retained include mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2). These metrics are widely used in weather forecasting because they allow both physical interpretation of errors and direct comparison with existing work.

3.1. Quantitative Performance of the Proposed Model

For an input sequence length of 72 hours, the Wavelet-MHA-LSTM model achieves high overall performance, with an average MAE less than 1.3 all variables combined. The temperature at 2m is predicted with an average error of less than 0.4°C, which is particularly satisfactory given the diurnal and seasonal variability observed in the data. Relative humidity presents an MAE of the order of 2 to 3%, while the wind speed at 10m displays an MAE close to 1km/h. When the sequence

length is increased to 120 hours, performance remains remarkably stable (See the results of the metrics in Table A1, which are almost the same as when the sequence length is 72 hours). This stability indicates that the model does not suffer from degradation linked to the increase in the time window, a phenomenon frequently observed with classic LSTM architectures. The coefficient R^2 remains high for temperature and humidity, confirming the capacity of the model to explain a large part of the observed variance. This robustness with respect to the length of the sequence suggests that the proposed architecture is capable of efficiently capturing medium and long term temporal dependencies, while limiting the effects of overlearning or information dilution. Table 1 summarizes the performance of the proposed hybrid model.

Table 1. Average performances (72h horizon).

Metric	Value	Mean Value
MAE_temperature_2m (°C)	0.352521	
MAE_relative_humidity_2m (%)	2.186118	1.210171
MAE_wind_speed_10m (km/h)	1.091873	
RMSE_temperature_2m (°C)	0.516662	
RMSE_relative_humidity_2m (%)	3.108643	1.713185
RMSE_wind_speed_10m (km/h)	1.514250	
R^2 _temperature_2m (°C)	0.978361	
R^2 _relative_humidity_2m (%)	0.984563	0.944284
R^2 _wind_speed_10m (km/h)	0.869928	

3.2. Qualitative Analysis of Temporal Predictions

Figures 3–5 compare the observed and predicted weekly values of the target variables during the first 500 time steps of the test set (equivalent to July 2024 to December 2024) for the six Togolese cities in this study. These curves highlight a strong temporal consistency between the ERA5 measurements and the predictions obtained in this work. For temperature at 2m, the model accurately reproduces diurnal cycles, characterized by regular and well-marked variations, as well as slower fluctuations associated with changes in air mass. Relative humidity, a variable that is notoriously more unstable and strongly correlated with local conditions, presents abrupt variations, particularly during day–night transitions. The model nevertheless manages to follow these fluctuations with satisfactory precision, although deviations remain during extreme changes, particularly when humidity approaches saturation values. Wind speed is the most difficult variable to predict, due to atmospheric turbulence, shear effects and local influences not observed in the data. Despite these intrinsic difficulties, the model correctly reproduces the dominant wind regimes and medium frequency variations. The largest errors appear mainly during sudden bursts, which is consistent with the deterministic nature of the model and the absence of explicit uncertainty modeling.

Table 2 presents a quantitative comparison of the average performances obtained by the three models considered for a 72 hour forecast. These results confirm the significant contribution of the hybrid architecture, with a reduction in the MAE of almost 32% compared to the standard LSTM and 19% compared to the CNN–LSTM.

Table 2. Comparison of average performances for the 72-hour forecasting horizon.

Model	MAE	RMSE	R^2
LSTM	1.78	2.45	0.82
CNN-LSTM	1.49	2.08	0.87
Wavelet-MHA-LSTM	1.21	2.01	0.94

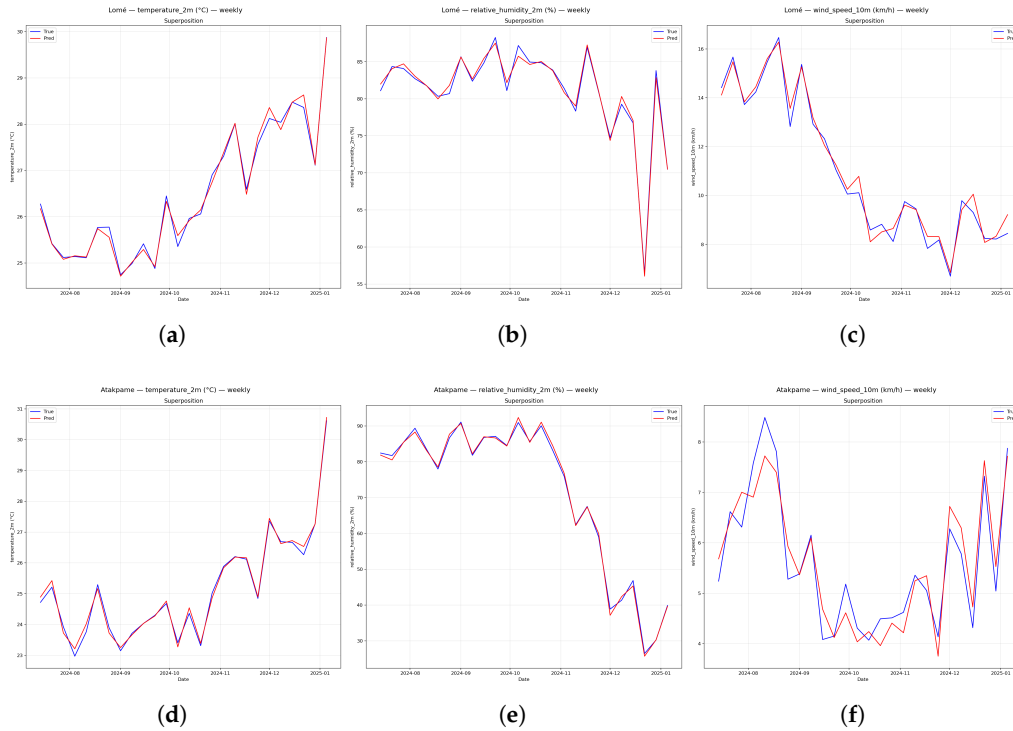


Figure 3. Comparative curves of weekly variations (July to December 2024) in temperature, relative humidity, and wind speed between the results obtained by ERA5 (True) and the predictions in this study (pred): case of the cities of Lomé and Atakpame (Togo).

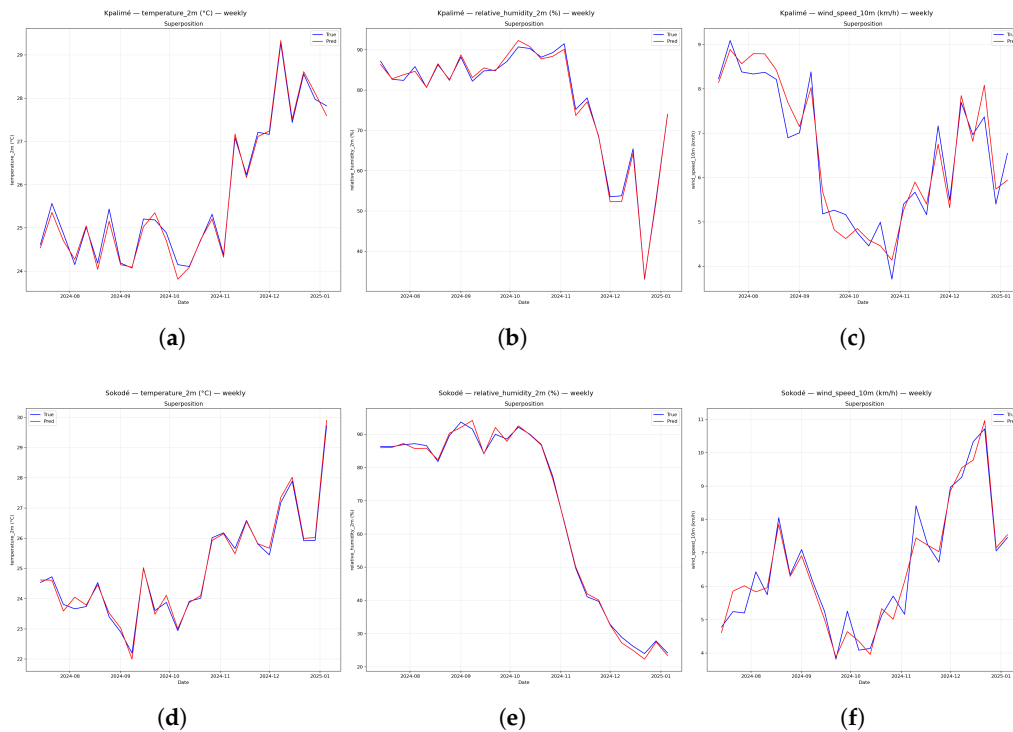


Figure 4. Comparative curves of weekly variations (July to December 2024) in temperature, relative humidity, and wind speed between the results obtained by ERA5 (True) and the predictions in this study (pred): case of the cities of Kpalimé and Sokodé (Togo).

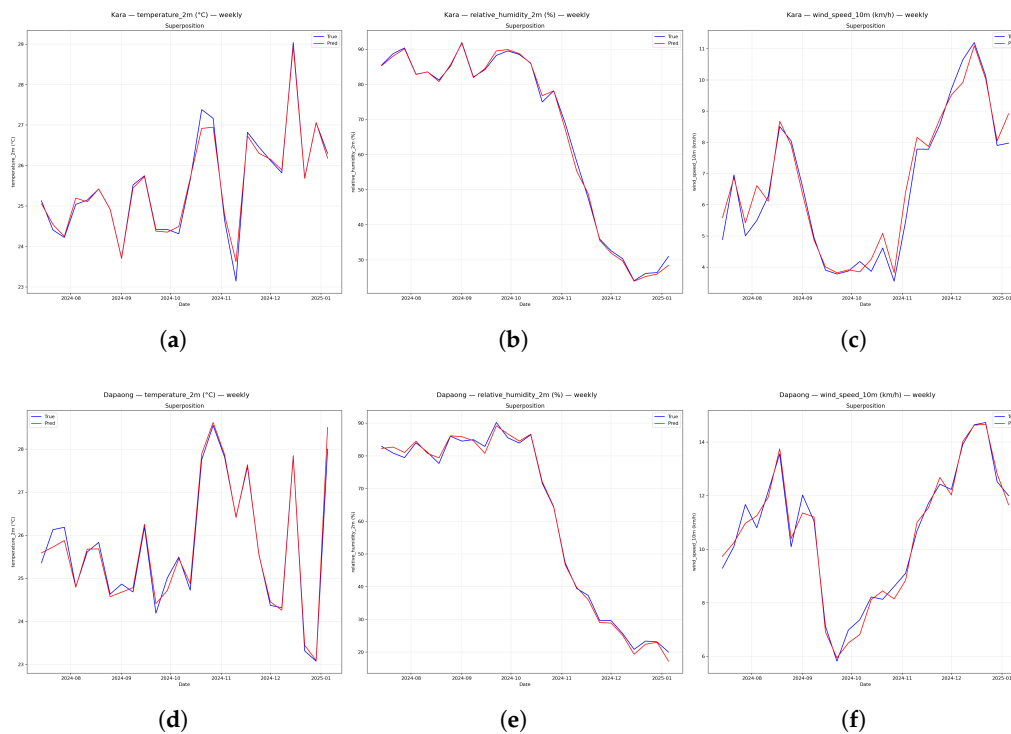


Figure 5. Comparative curves of weekly variations (July to December 2024) in temperature, relative humidity, and wind speed between the results obtained by ERA5 (True) and the predictions in this study (pred): case of the cities of Kara and Dapaong (Togo).

4. Discussion

The results obtained with the proposed SWT–Multi-Head Attention–LSTM architecture demonstrate the relevance of combining explicit multi-scale signal decomposition with adaptive feature weighting and sequential temporal modeling for short-term meteorological forecasting. The hybrid structure allows the model to simultaneously capture slow atmospheric trends (seasonal components), medium-scale variability (synoptic fluctuations), and high-frequency oscillations (diurnal cycles and short-lived wind gusts). In tropical environments such as Togo, where rapid transitions between dry and rainy conditions are common, this multi-scale capability is particularly important.

The integration of Stationary Wavelet Transform (SWT) contributes significantly to performance stability. Unlike classical time-series models that operate directly on raw sequences, the SWT layer exposes structured frequency components that help disentangle long-term variability from transient disturbances. This is especially beneficial for wind speed and relative humidity, which exhibit intermittent spikes and nonlinear fluctuations. Furthermore, the Multi-Head Attention mechanism enhances interpretability by dynamically adjusting the relative importance of different scales and features, enabling adaptive weighting rather than static feature fusion. However, despite its promising performance, several limitations must be acknowledged. First, the quality and representativeness of the dataset remain critical constraints. The model relies on historical meteorological data that may originate from reanalysis products or global datasets. While these sources provide broad coverage, they may not fully reflect microclimatic variations specific to different regions of Togo (e.g., coastal Lomé vs. northern savanna zones such as Dapaong). High-resolution in situ measurements from national meteorological stations would significantly improve model reliability. The inclusion of long-term station-based observations from the Togolese meteorological agency (covering multiple decades and diverse ecological zones) would allow better calibration, improved generalization, and more realistic extreme-event modeling. Second, dataset size and temporal depth directly impact the robustness of deep learning models. Although the current framework supports several years of hourly data, deeper temporal coverage would enhance seasonal learning, particularly for rare extreme

events such as intense rainfall bursts or Harmattan wind episodes. A sufficiently large dataset is essential to stabilize attention weights and prevent overfitting, especially when wavelet decomposition introduces additional feature dimensions. Third, the current implementation focuses on three target variables: temperature, relative humidity, and wind speed. While these are fundamental atmospheric indicators, meteorological dynamics are inherently multivariate and coupled. Incorporating additional predictors such as precipitation, surface pressure, solar radiation (shortwave and diffuse components), cloud cover stratification, evapotranspiration, soil moisture, and atmospheric stability indices could improve predictive accuracy. These variables provide richer physical context and allow the model to better capture energy balance and moisture transport mechanisms. In particular, radiation variables are crucial for temperature forecasting, and pressure gradients strongly influence wind behavior. Another limitation concerns spatial dependency modeling. The present architecture treats cities independently except through embedding representation. However, meteorological processes are spatially correlated. Advection mechanisms transport humidity and temperature across regions. A spatio-temporal extension incorporating graph neural networks or attention across cities could better model inter-city dependencies, especially between southern and northern climatic zones of Togo. Computational complexity also represents a constraint. Although the reduced grid-search and multi-node SLURM parallelization make training feasible within a few hours, wavelet decomposition and attention mechanisms increase memory consumption and training time. For operational deployment, model compression or pruning strategies may be required. Finally, the model assumes stationarity within training segments. Climate variability and long-term climate change trends may introduce distribution shifts that degrade performance over time. Periodic retraining and online adaptation strategies would therefore be necessary in a real-world forecasting system.

5. Conclusions

In this article, an innovative hybrid model has been proposed for multivariate weather forecasting from hourly time series. The model combines wavelet decomposition, multi-head attention mechanism and LSTM network to effectively capture multi-scale temporal dependencies and complex nonlinearities present in meteorological data. The results demonstrate that the proposed architecture significantly outperforms benchmark models such as classical LSTM and CNN-LSTM, both in terms of accuracy and stability, for the prediction of temperature, relative humidity and wind speed. The robustness observed with respect to the length of the input sequence confirms the ability of the model to generalize over extended time horizons. Despite these promising performances, several avenues for improvement can be considered. The explicit integration of spatial dependencies between different weather stations, via graph networks or spatio-temporal Transformers, constitutes a natural extension of this work. In addition, the adoption of probabilistic or Bayesian approaches would make it possible to quantify the uncertainty associated with forecasts, a crucial aspect for operational applications. Finally, the application of the model to other climatic contexts and to data at higher temporal or spatial resolution would make it possible to assess its genericity and its potential for large-scale weather forecasting systems.

Author Contributions: Apeke Sena and Gbafa Senanou designed the study and defined the methodology. Kpogo-Nuwoklo K. Agbéko brought expertise in the field. Sena Apeke developed the code of the proposed model, wrote the Materials and Methods section. Gbafa Senanou and Wotodzo Kokou Felicien proposed the introduction. Kpogo-Nuwoklo K. Agbéko and Sena Apeke did the rest. The whole group reread the work and everyone made corrections. Professor Agboka Komi and Professor Ouro-Djobo led the entire project.

Funding: This research received no external funding.

Data Availability Statement: C3S: ERA5 hourly data on single levels from 1940 to present, 2018. <https://doi.org/10.24381/CDS.ADBB2D47>.

Acknowledgments: We thank the European Union Program, Copernicus, Climate Change Service and ECMWF, for the fundraising work they are doing to put in place interesting models and databases to enable the evolution

of research in general around the world. We also thank Mrs. GBEDEVI Yawa Jacqueline who did an internship on data from ERA5.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Tables and Figures

Table A1. Average performances (120h horizon).

Metric	Value	Mean Value
MAE_temperature_2m (°C)	0.351253	
MAE_relative_humidity_2m (%)	2.198971	1.216478
MAE_wind_speed_10m (km/h)	1.099211	
RMSE_temperature_2m (°C)	0.518849	
RMSE_relative_humidity_2m (%)	3.095541	1.710990
RMSE_wind_speed_10m (km/h)	1.518582	
R^2 _temperature_2m (°C)	0.977808	
R^2 _relative_humidity_2m (%)	0.984651	0.944271
R^2 _wind_speed_10m (km/h)	0.870356	

Table A2. Target variables and primary atmospheric explanatory features.

Variable Name	Description	Unit
Target Variables		
temperature_2m	Air temperature measured at 2 meters above ground level	°C
relative_humidity_2m	Relative humidity at 2 meters above ground	%
wind_speed_10m	Wind speed measured at 10 meters above ground	km/h
Atmospheric Explanatory Variables		
dew_point_2m	Dew point temperature at 2 meters	°C
apparent_temperature	Perceived temperature combining air temperature, humidity, and wind	°C
pressure_msl	Atmospheric pressure reduced to mean sea level	hPa
cloud_cover	Total cloud cover fraction	%
precipitation	Instantaneous precipitation amount	mm
vapour_pressure_deficit	Vapor pressure deficit (indicator of atmospheric dryness)	kPa
wind_u	Zonal wind component (east–west direction)	km/h
wind_v	Meridional wind component (north–south direction)	km/h
absolute_humidity	Absolute humidity of the air	g/m ³
heat_index	Heat index combining temperature and humidity effects	°C
dewpoint_spread	Difference between air temperature and dew point temperature	°C

Table A3. Radiative, temporal, and soil-related input variables.

Variable Name	Description	Unit
Radiative Variables		
shortwave_radiation_instant	Incoming global shortwave solar radiation	W/m ²
direct_normal_irradiance_instant	Direct normal solar irradiance	W/m ²
diffuse_radiation_instant	Diffuse solar radiation component	W/m ²
terrestrial_radiation_instant	Outgoing terrestrial (infrared) radiation	W/m ²
Temporal Encoded Variables		
hour_sin, hour_cos	Sinusoidal encoding of hourly cycle (diurnal periodicity)	–
month_sin, month_cos	Cyclical encoding of annual seasonality (month)	–
dayofweek_sin, dayofweek_cos	Cyclical encoding of weekly periodicity	–
Soil and Hydrological Variables		
soil_temperature_0_to_7cm	Soil temperature at 0–7 cm depth	°C
soil_moisture_0_to_7cm	Volumetric soil moisture at 0–7 cm depth	m ³ /m ³
et0_fao_evapotranspiration	Reference evapotranspiration (FAO Penman–Monteith)	mm

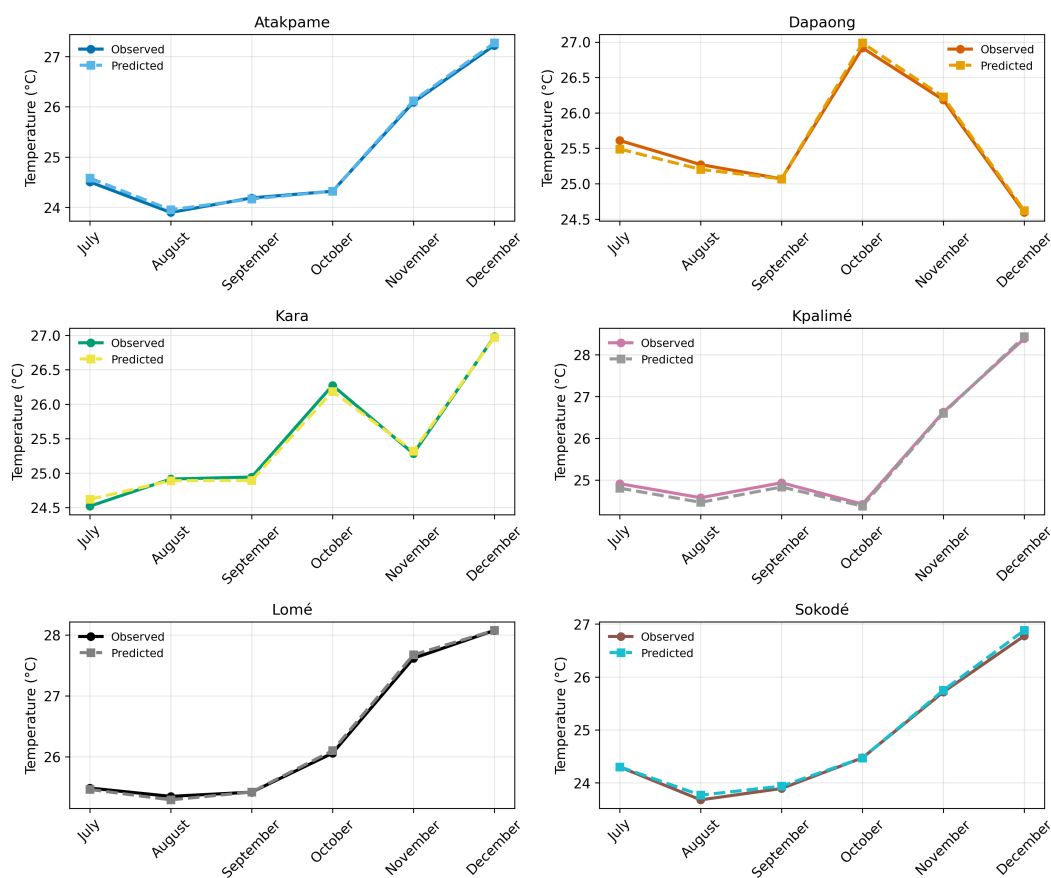


Figure A1. Comparison of observed and predicted average values of temperature at 2m over the period from July to December 2024 in Togo, case for a length prediction window 72.

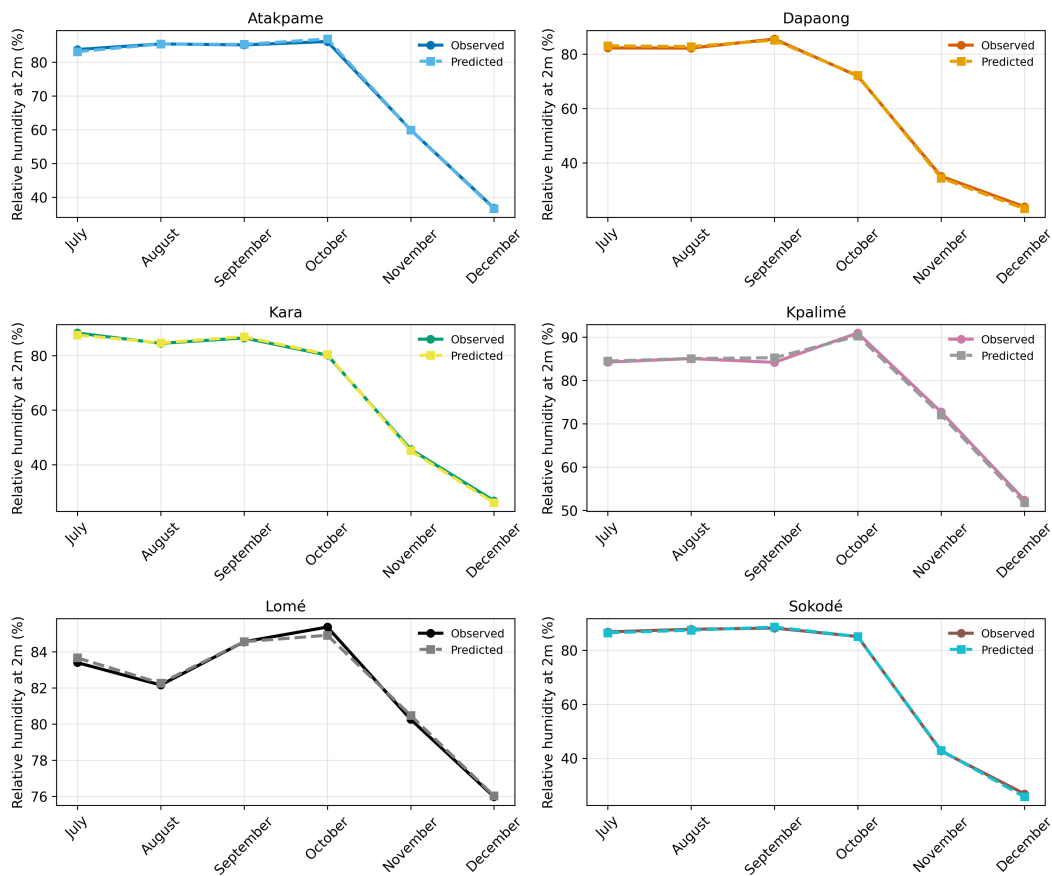


Figure A2. Comparison of observed and predicted average values of relative humidity at 2m over the period from July to December 2024 in Togo, case for a length prediction window 72.

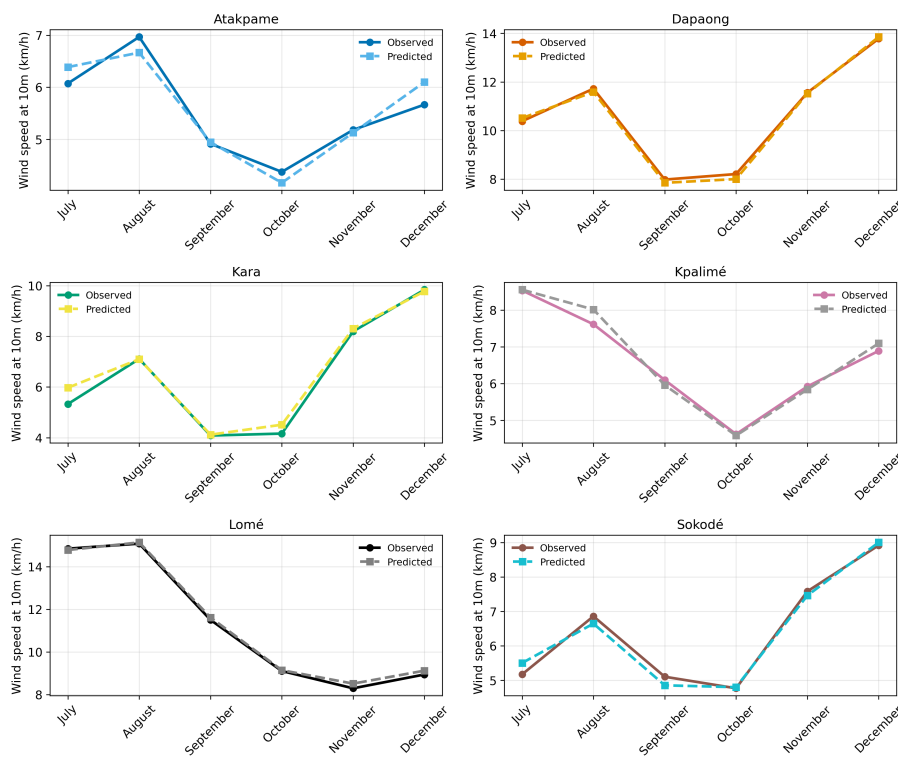


Figure A3. Comparison of observed and predicted average wind speed values at 10m over the period from July to December 2024 in Togo, case for a length prediction window 72.

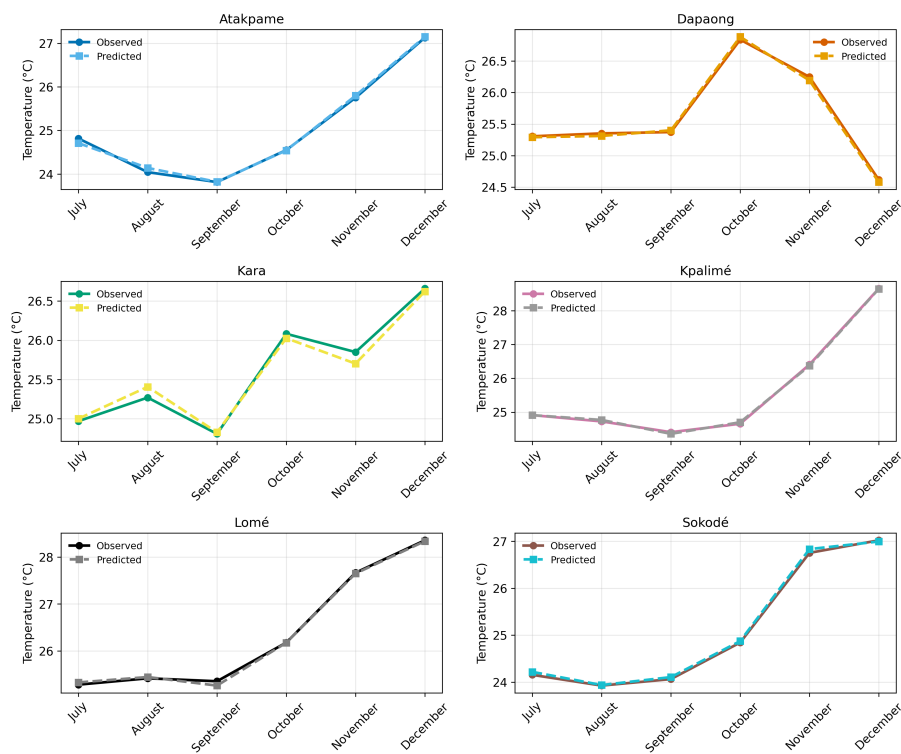


Figure A4. Comparison of observed and predicted average values of temperature at 2m over the period from July to December 2024 in Togo, case for a length prediction window 120.

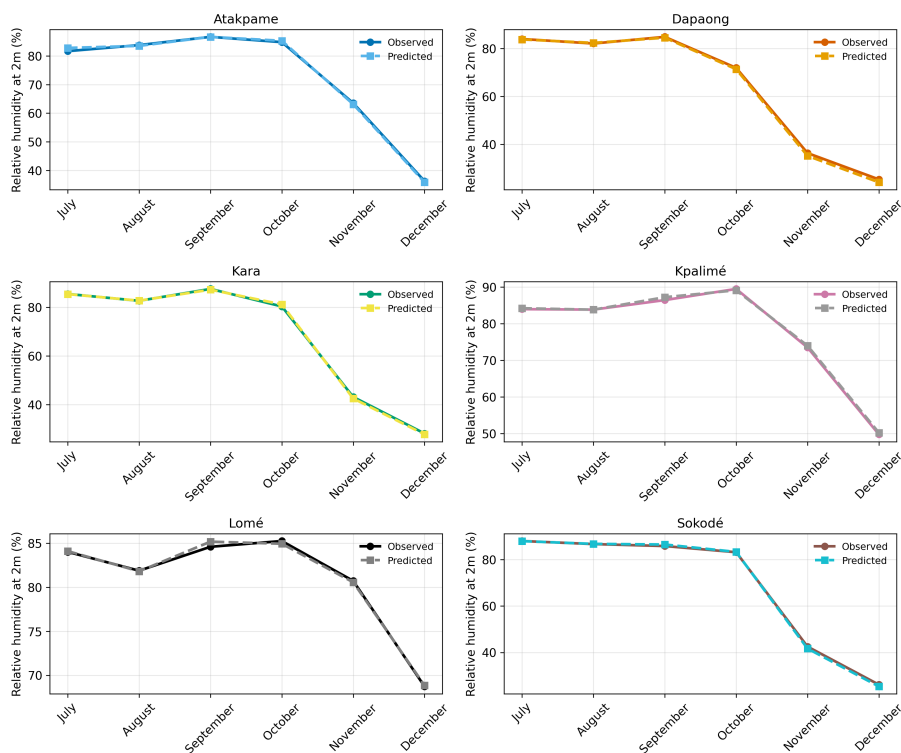


Figure A5. Comparison of observed and predicted average values of relative humidity at 2m over the period from July to December 2024 in Togo, case for a length prediction window 120.

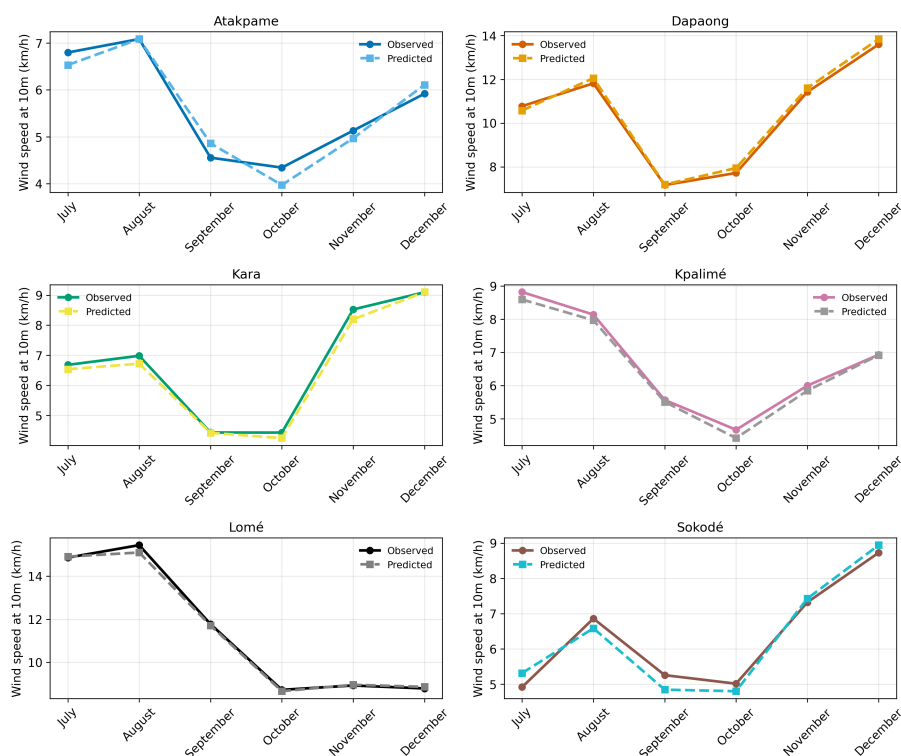


Figure A6. Comparison of observed and predicted average wind speed values at 10m over the period from July to December 2024 in Togo, case for a length prediction window 120.

References

1. Stull, R.B. *An Introduction to Boundary Layer Meteorology*; Springer: Dordrecht, 1988. <https://doi.org/10.1007/978-94-009-3027-8>.
2. Wallace, John M., H.P.V. *Atmospheric Science: An Introductory Survey*, 2 ed.; Elsevier: Amsterdam, 2006.
3. Monteith, J.L.; Unsworth, M.H. *Principles of Environmental Physics: Plants, Animals, and the Atmosphere*, 4 ed.; Academic Press: Oxford, 2013.
4. Sultan, B.; Janicot, S. The West African Monsoon Dynamics. Part II: The “Preonset” and “Onset” of the Summer Monsoon. *Journal of Climate* **2003**, *16*, 3407–3427. [https://doi.org/10.1175/1520-0442\(2003\)016<3407:TWAMDP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3407:TWAMDP>2.0.CO;2).
5. Nicholson, S.E. The West African Sahel: A Review of Recent Studies on the Rainfall Regime and Its Interannual Variability. *ISRN Meteorology* **2013**, pp. 1–32. <https://doi.org/10.1155/2013/453521>.
6. Sylla, M.B.; Nikiema, P.M.; Gibba, P.; Kebe, I.; Klutse, N.A.B. Projected Changes in the Annual Cycle of High-Intensity Precipitation Events over West Africa for the late twenty-first century. *Journal of Climate* **2015**, *28*, 6475–6491. <https://doi.org/10.1175/JCLI-D-14-00654.1>.
7. Guichard, F.; Kergoat, L.; Mougin, E.; et al.. Modelling the diurnal cycle of deep precipitating convection over land with global and regional models. *Quarterly Journal of the Royal Meteorological Society* **2009**, *135*, 832–858. <https://doi.org/10.1002/qj.406>.
8. FAO. *The Future of Food and Agriculture: Trends and Challenges*; Food and Agriculture Organization of the United Nations: Rome, 2017.
9. Gasparrini, A.; Guo, Y.; Hashizume, M.; et al.. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet* **2015**, *386*, 369–375. [https://doi.org/10.1016/S0140-6736\(14\)62114-0](https://doi.org/10.1016/S0140-6736(14)62114-0).
10. World Health Organization. *COP24 Special Report: Health and Climate Change*; World Health Organization: Geneva, 2018.
11. Lorenz, E.N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **1963**, *20*, 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2).
12. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*; Cambridge University Press: Cambridge, 2003. <https://doi.org/10.1017/CBO9780511802270>.

13. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, revised ed.; Holden-Day: San Francisco, 1976.
14. Chatfield, C. *Time-Series Forecasting*; Chapman and Hall/CRC: Boca Raton, 2000.
15. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*; Academic Press, 2019.
16. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3 ed.; OTexts: Melbourne, 2021. Open-access online textbook.
17. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5–32. <https://doi.org/10.1023/A:1010933404324>.
18. Cortes, C.; Vapnik, V. Support-Vector Networks. *Machine Learning* **1995**, *20*, 273–297. <https://doi.org/10.1007/BF00994018>.
19. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, *29*, 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
20. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, 2006.
21. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9*, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
22. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.c. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 802–810.
23. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
24. Ngoungue Langue, C.G.; Lavaysse, C.; Flamant, C. Subseasonal forecasts of heat waves in West African cities. *Natural Hazards and Earth System Sciences* **2025**, *25*, 147–168. <https://doi.org/10.5194/nhess-25-147-2025>.
25. Awe, O.; Dias, R.; Ajetunmobi, T.; Ayeni, O.; Oyanameh, O.; Agunloye, O.K. *Time Series Forecasting of Seasonal Non-stationary Climate Data: A Comparative Study*; 2023; pp. 335–350. https://doi.org/10.1007/978-3-031-41352-0_17.
26. C3S. ERA5 hourly data on single levels from 1940 to present, 2018. <https://doi.org/10.24381/CDS.ADBB2D47>.
27. Direction Générale Météorologie Nationale, M.d.I.e.d.T. PLAN D’ACTION NATIONAL POUR LA MISE EN PLACE DU CADRE NATIONAL POUR LES SERVICES CLIMATOLOGIQUES (CNSC) AU TOGO. Technical report.
28. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press, 2000. <https://doi.org/10.1017/CBO9780511841040>.
29. Mallat, S. *A Wavelet Tour of Signal Processing*; Academic Press, 1999. <https://doi.org/10.1016/B978-012466606-1/50000-6>.
30. Qin, Y.; Song, D.; Chen, H.; Cheng, W.; Jiang, G.; Cottrell, G. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. In Proceedings of the Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), 2017. <https://doi.org/10.24963/ijcai.2017/366>.
31. Lim, B.; Arık, S.O.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *International Journal of Forecasting* **2021**, *37*, 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
32. Bandara, K.; Bergmeir, C.; Smyl, S. LSTM-Based Ensemble for Time Series Forecasting. *International Journal of Forecasting* **2020**, *36*, 86–99. <https://doi.org/10.1016/j.ijforecast.2019.01.011>.

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