Science Journal of Energy Engineering
2026; 14(3): 65-82
http://www.sciencepublishinggroup.com/j/sjee
doi: 10.11648/j.sjee.20261403.11

Comparative Analysis of LSTM and Hybrid Attention-Based Architectures for Short-Term Electricity Consumption Forecasting: A Case Study in Togo

Sena Apeke[1,2,*], Yao Bokovi[1,2], Kodjovi Senanou Gbafa[3], Komlan Kolegain[2], Sanoussi Ouro-Djobo[2,4]

[1]Engineering Science Research Laboratory, National Polytechnic School, Lome, Togo
[2]Regional Center of Excellence for Electricity Control, Solar Energy Laboratory, University of Lome, Lome, Togo
[3]Laboratory of Structures and Materials Mechanics, National Polytechnic School, Lome, Togo
[4]Solar Energy Laboratory, Department of Physics, Universite of Lome, Lome, Togo

Email address:
sapeke@univ-lome.tg (Sena Apeke), bokoviyao@gmail.com (Yao Bokovi), gbafafelix@gmail.com (Kodjovi Senanou Gbafa), yvescole@gmail.com (Komlan Kolegain), odsanoussi@gmail.com (Sanoussi Ouro-Djobo)

To cite this article:
Sena Apeke, Yao Bokovi, Kodjovi Senanou Gbafa, Komlan Kolegain, Sanoussi Ouro-Djobo. (2026). Comparative Analysis of LSTM and Hybrid Attention-Based Architectures for Short-Term Electricity Consumption Forecasting: A Case Study in Togo. *Science Journal of Energy Engineering, 14*(3), 65-82. https://doi.org/10.11648/j.sjee.20261403.11

Received: 18 June 2026; **Accepted:** 6 July 2026 **Published:** 21 August 2026

Abstract: TAccurate short-term electricity consumption forecasting is essential for operational planning, reserve allocation, and energy management in modern power systems. This study investigates the performance of recurrent and hybrid attention-based deep learning architectures for short-term electricity consumption forecasting, a sase study in Togo. The proposed framework integrates electricity consumption, meteorological, demographic, and temporal information in order to capture both intrinsic temporal dependencies and exogenous influences affecting electricity demand. Four forecasting architectures were evaluated using a weekly temporal window of 168 hours: LSTM-only, LSTM-decoder, LSTM-attention, and a hybrid LSTM–Multi-Head Attention–LSTM model. The experiments included multi-seed evaluation, ablation study, robustness analysis, and statistical comparison using the Wilcoxon signed-rank test. The results show that all models achieved extremely high forecasting accuracy, with coefficients of determination exceeding 0.9998 and MAPE values below 0.004%. The hybrid architectures slightly improved average forecasting performance, while the standalone LSTM model remained highly competitive. The robustness analysis revealed strong sensitivity to noisy inputs but moderate degradation under missing-data conditions. Statistical analysis indicated that the performance differences between architectures were not statistically significant at the 5% level. Overall, the study demonstrates that recurrent deep learning architectures provide reliable and operationally relevant solutions for electricity consumption forecasting in highly structured energy demand series.

Keywords: Model, LSTM, Multi-Head Attention, Energy, Prediction, Consumption

1. Introduction

Electricity is an essential pillar of economic, social, and technological development. It supports industrial growth and helps improve people's living conditions. For developing countries, the availability of reliable and accessible electricity is all the more crucial as it determines economic competitiveness, investment attractiveness, and social well-being. System operators rely on accurate forecasts to plan production units, allocate spinning reserves, manage maintenance, optimize cross-border imports, and mitigate operational risks associated with underestimating or overestimating demand [7, 8]. In Togo, demand for electricity is growing steadily, driven by several factors: rapid urbanization, population growth, modernization, and productive activities. According to CEET (Compagnie Énergie Électrique du Togo), electricity consumption has risen sharply over the past ten years, reflecting both a growing need for energy and increased pressure on the distribution network. However, managing this consumption is marked by persistent challenges such as technical and non-technical losses, seasonal fluctuations, inadequate production planning, and constraints related to importing electricity to meet domestic demand. These constraints make demand forecasting particularly strategic in order to reduce imbalances between supply and demand. Electricity consumption forecasting has undergone major changes over the past ten years. Early work relied on empirical models: AutoRegressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Double SARIMA (DSARIMA) [11, 12], then machine learning approaches became prevalent, before the emergence of deep learning and Transformer models. Dinata et al. [1] proposed a bi-seasonal ARIMA model for load demand, but the management of nonlinear relationships is a limitation. Nguyen et al. [4], Ajlouni, Sameh A. [5] proposed an online variant of the seasonal autoregressive integrated moving average (SARIMA) to predict electrical load sequentially, but their model has the same limitations as that proposed by Dinata et al., capturing weak nonlinearity. In response, machine learning (ML) approaches such as support vector regression and ensemble trees have been adopted to better capture nonlinear mappings, often improving accuracy but still relying heavily on feature engineering and sometimes exhibiting limited extrapolation robustness in the face of structural change [15]. Deep learning has significantly advanced short-term load forecasting (STLF) by learning representations directly from sequential data. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) models, capture temporal dependencies via closed memory and have repeatedly demonstrated strong performance for load forecasting tasks [13, 16]. However, LSTM family models can still struggle to model long-range dependencies and multi-scale periodicities (daily/weekly/annual) in long sequences without careful design. Attention mechanisms and Transformer-like architectures address this limitation by allowing direct interactions between time steps and learning to focus on informative historical segments [17, 18]. Recent work has shown that multi-head attention can improve forecasting performance by capturing global trends and complex seasonal structures, especially when combined with convolutional or recurrent components [19, 20]. In Togo, several studies have examined the modeling and forecasting of electricity consumption with varying scopes. A short term study using multiple linear regression employed calendar and lag characteristics to predict demand and reported low MAPE with high R^2 over the period examined, but the approach remains limited by linear assumptions and modest capacity for nonlinear temporal interactions [27]. For longer term planning, system dynamics approaches have incorporated socioeconomic indicators such as population and GDP to project residential consumption in Lomé toward 2030-2050, highlighting population growth as a central driver [28]. More recently, studies focusing on wetland contexts (with Togo as a case study) have evaluated meteorological predictors (temperature, humidity, precipitation, wind, sunshine duration) with models such as Extreme Gradient Boosting (XGBoost), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Recurrent Neural Network (RNN), concluding that RNN type models can provide higher R^2 values than the other methods tested, while also highlighting the role of humidity and precipitation correlations [23]. In parallel, recent contributions from Togo have explored ML and hybrid deep learning for load forecasting in microgrids or related contexts, reflecting a growing interest in data-driven demand modeling [14]. Collectively, these studies demonstrate the relevance of forecasting for the Togolese electricity system, but they also illustrate common limitations: (i) reliance on either consumption signals alone or a limited set of exogenous variables, (ii) limited modeling of multi-scale periodicity (particularly weekly cycles), and (iii) poor integration of demographic dynamics alongside weather conditions and consumption history.

Electricity demand is a socio-technical outcome determined by both short-term factors and longer-term structural factors. In the short term, weather conditions such as temperature and humidity influence cooling/heating needs and thus strongly modulate demand, with lagged or cumulative effects that are not always captured by simple instantaneous predictors [7, 24, 25]. In the medium and long term, demographic trends (population growth, urbanization) and economic development alter the reference level and elasticity of demand. Demand forecasting analyses consistently identify socio-economic variables (population, GDP, per capita usage) as key determinants for long-term projections [15, 28]. However, many modern deep learning STLF pipelines are built primarily around historical load (and possibly calendar variables), partly because exogenous data may be missing, poorly aligned, or more difficult to operationalize [8, 10]. This gap is particularly important in emerging networks where climate sensitivity and population growth are pronounced and where structural changes can degrade purely autoregressive models.

To overcome these limitations, this study aims to provide realistic electricity consumption forecasts for Togo by jointly modeling three complementary sources of information: (1) historical electricity consumption, (2) meteorological variables, and (3) demographic indicators. We propose a hybrid deep learning pipeline that combines recurrent sequence modeling and attention-based long-term dependency learning. Specifically, the model uses an LSTM-based encoder to capture local temporal dynamics and nonlinear interactions between heterogeneous inputs, a multi-head residual attention block to represent structured weekly seasonality and global trends, and a final LSTM step (or decoder) to refine the representation towards accurate predictions with a lead time. The design explicitly exploits weekly seasonality with a period of $s = 168$ hours, aligning the input window with typical operational planning cycles and weekly usage patterns.

Main contributions

The main contributions of this work are summarized as follows:

1) A multimodal forecasting strategy is developed by integrating electricity consumption data with meteorological and demographic information, allowing the models to capture both temporal dynamics and exogenous influences affecting electricity demand.
2) A statistical comparison based on the Wilcoxon signed-rank test is performed to evaluate the significance of the performance differences between the forecasting architectures.
3) An ablation study is carried out to quantify the contribution of each component of the hybrid architecture, including the recurrent encoder, the attention mechanism, and the decoder structure.
4) A robustness analysis is conducted under noisy and partially missing data conditions to assess the stability and generalization capability of the proposed forecasting approaches.
5) From an energy management perspective, the proposed framework provides practical insights for improving short-term load forecasting, operational planning, and decision-making in developing electrical power systems such as those of West Africa.

2. Materials and Methods

2.1. Data Collection and Description

This study relies on a multimodal dataset combining electricity consumption, meteorological, demographic, and calendar-related information in order to improve short-term electricity consumption forecasting. The integration of heterogeneous sources aims to capture both intrinsic temporal dynamics and exogenous factors influencing electricity demand.

Electricity consumption data were provided by the *Togo Electric Power Company (CEET)* and cover the period from 2014 to 2018 with an hourly temporal resolution. Each observation corresponds to the national electricity consumption measured in MWh and represented by the target variable denoted as `CONSOMMATION_TOTALE`. The dataset also includes temporal descriptors such as hour, day, month, and year. A preliminary analysis confirmed the temporal consistency of the database, with no duplicate timestamps and only negligible missing values. Figure 1 presents the structure and a brief overview of the data used in this study.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
DATE HEURES	MOIS	ANNEE	CONSOMMATION	TEMPERATURE	HUMIDI	POINT ROSEE	PRECIPITATION	VENT	POPULATION	DENSITE	URBANISATION	HOLIDAY	WEEK END
2014-01-01 00:00:00	1	2014	219.9	24.9	73.5	18.625	0.0	11.675	7468625	137.40067546106692	39.53802530622082	1	0
2014-01-01 01:00:00	1	2014	218.32	24.425	75.5	18.775	0.0	11.125	7468626	137.4007050846541	39.538035964871575	1	0
2014-01-01 02:00:00	1	2014	213.88	24.125	75.75	18.7	0.0	10.8	7468628	137.40076312270247	39.538056847126136	1	0
2014-01-01 03:00:00	1	2014	210.83999999999997	23.85	77.25	18.9	0.0	10.025	7468632	137.4008484154381	39.53808753569454	1	0
2014-01-01 04:00:00	1	2014	205.61999999999998	23.6	78.75	19.15	0.0	8.775	7468636	137.40095985042475	39.538127630319075	1	0
2014-01-01 05:00:00	1	2014	194.86	23.375	80.5	19.4	0.025	7.75	7468642	137.4010963606318	39.538176747079085	1	0
2014-01-01 06:00:00	1	2014	174.94	23.15	83.5	19.925	0.09999999999999999	6.525	7468648	137.40125692258087	39.53823451772415	1	0
2014-01-01 07:00:00	1	2014	164.96	23.525	83.75	20.575	0.025	6.875	7468655	137.40144055456818	39.53830058903446	1	0
2014-01-01 08:00:00	1	2014	159.08	25.299999999999997	76.75	20.65	0.05	6.95	7468663	137.40164631495952	39.538374622207364	1	0
2014-01-01 09:00:00	1	2014	160.95999999999998	26.95	73.5	21.7	0.0	6.7749999999999995	7468672	137.40187330055474	39.538456292268876	1	0
2014-01-01 10:00:00	1	2014	161.5	28.6	67.5	21.825	0.0	8.049999999999999	7468682	137.40212064501898	39.53854528750925	1	0
2014-01-01 11:00:00	1	2014	165.4	29.75	61.25	21.05	0.0	8.5	7468693	137.40238751737797	39.53864130894159	1	0
2014-01-01 12:00:00	1	2014	167.60000000000002	30.675	55.75	20.05	0.05	9.2	7468704	137.40267312057483	39.53874406978257	1	0
2014-01-01 13:00:00	1	2014	167.0	30.625	54.75	19.25	0.15	9.5	7468716	137.40297669008552	39.538853294954265	1	0
2014-01-01 14:00:00	1	2014	168.5	30.7	53.75	18.8	0.225	9.025	7468729	137.40329749259112	39.538968720606455	1	0
2014-01-01 15:00:00	1	2014	157.8	30.775	53.0	18.525	0.225	9.35	7468742	137.4036348247041	39.5390900936583	1	0
2014-01-01 16:00:00	1	2014	152.18	30.45	53.5	18.275	0.025	8.9	7468756	137.4039880117466	39.5392171713588	1	0
2014-01-01 17:00:00	1	2014	149.18	29.5	57.25	18.775	0.09999999999999999	8.649999999999999	7468771	137.40435640657864	39.53934972086514	1	0
2014-01-01 18:00:00	1	2014	367.64	28.325	62.25	19.375	0.0	8.325	7468786	137.40473938847413	39.53948751883831	1	0
2014-01-01 19:00:00	1	2014	435.55999999999995	27.55	66.25	19.849999999999998	0.0	7.3999999999999995	7468802	137.4051363620427	39.53963035105517	1	0
2014-01-01 20:00:00	1	2014	455.6	26.75	70.25	20.175	0.0	6.9	7468818	137.40554675619563	39.5397780120364	1	0
2014-01-01 21:00:00	1	2014	458.8	26.05	73.75	20.475	0.0	6.15	7468835	137.40597002315405	39.53993030468957	1	0

Figure 1. *Structure and overview of the multimodal dataset used in this study.*

Figure 2 illustrates the weekly electricity consumption profiles observed from 2014 to 2018. The curves reveal a highly structured and strongly autocorrelated time series characterized by recurrent daily and weekly consumption patterns.

Meteorological variables were collected using the public API *archive-api.open-meteo.com* [2]. The weather observations cover the same temporal interval (2014–2018) and originate from four representative climatic regions of Togo: Lomé, Anfoin, Atakpamé, and Kara. The collected variables include: Air temperature at 2 meters (°C), Relative humidity (%), Dew point (°C), Precipitation (mm) and Wind speed at 10 meters (m/s). To obtain a national meteorological representation synchronized with electricity consumption, hourly observations from the four cities were aggregated using temporal averaging.

Demographic indicators were extracted from the *World Bank* database [3]. Three variables were considered: Total population, Population density and Urbanization rate. Since demographic statistics are generally available at annual resolution, a temporal disaggregation procedure was applied to generate hourly demographic covariates aligned with electricity consumption data. The transformation process includes:

1) Linear interpolation from yearly to monthly and daily values,
2) Hourly temporal interpolation,
3) Addition of low-amplitude stochastic perturbations ($\pm 0.3\%$),
4) Exponential smoothing for temporal stabilization.

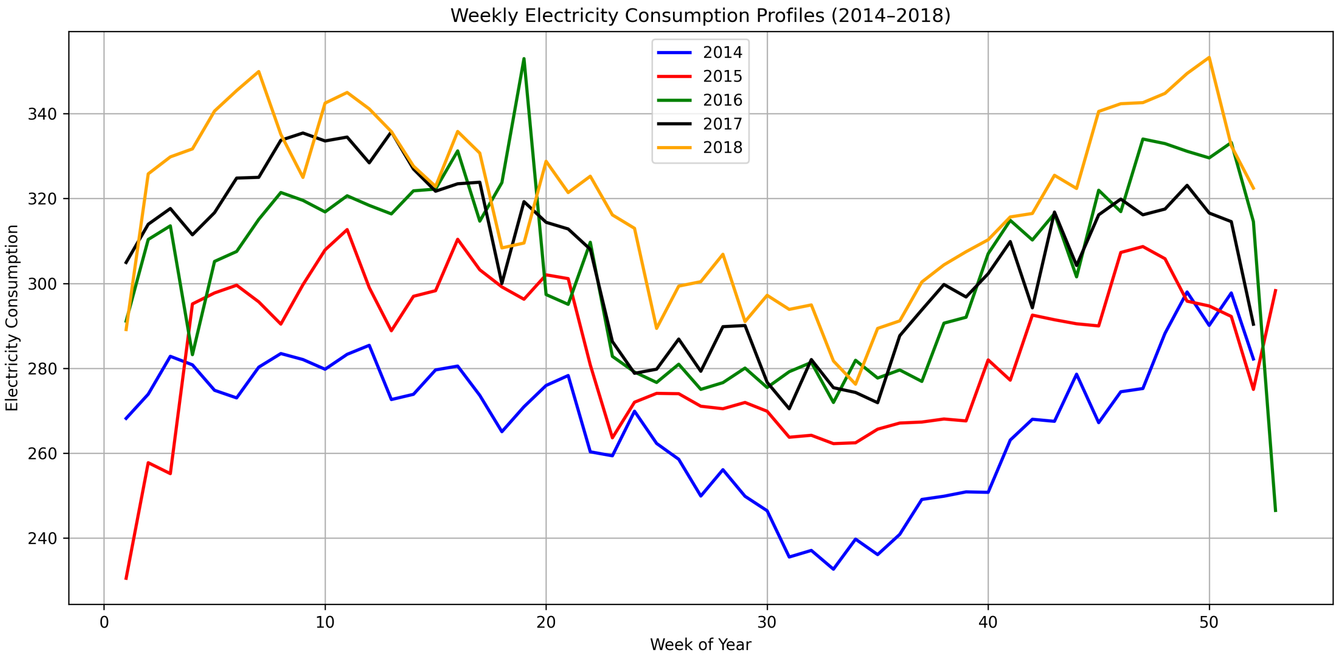

Figure 2. *Average weekly electricity consumption profiles from 2014 to 2018.*

This approach preserves long-term demographic trends while introducing moderate local variability compatible with short-term forecasting applications. Additional temporal features were constructed directly from timestamps to improve the representation of cyclical consumption behaviors. These variables include: Hour of day, day of week, month, weekend indicator, holiday indicator, sinusoidal temporal encodings. The sinusoidal encodings allow the model to preserve the cyclic continuity of temporal variables and improve periodic pattern learning. Table 1 summarizes the variables used in this study.

Table 1. *Description of variables used in the study.*

Variable	Unit	Description
DATE_TIME	Date-time	Timestamp of hourly observations.
CONSOMMATION_T	MWh	National electricity consumption.
TEMPERATURE	$^\circ$C	Average hourly air temperature.
HUMIDITY	%	Relative air humidity.
DEW_POINT	$^\circ$C	Dew point temperature.
PRECIPITATION	mm	Hourly precipitation amount.
WIND	m/s	Wind speed at 10 meters.
POPULATION	Inhabitants	Estimated national population.
DENSITY	inhabitants/km^2	Population density.
URBANIZATION	%	Urbanization rate.
HOLIDAY	(0/1)	National holiday indicator.
WEEKEND	(0/1)	Weekend indicator.

2.2. Ethics Approval and Consent to Participate

This study does not involve human participants, animals, or clinical data requiring ethical approval or informed consent. The research was conducted using aggregated electricity consumption, meteorological, demographic, and temporal data.

2.3. Problem Formulation

Let $\{y_t\}_{t=1}^{T}$ denote the hourly electricity consumption time series, where $y_t \in \mathbb{R}$, represents the national electricity consumption at time step t. Let $\mathbf{x}_t \in \mathbb{R}^d$ be the multivariate input vector composed of:

1) Historical electricity consumption,
2) Meteorological variables,
3) Demographic indicators,
4) Calendar-related variables.

Considering a temporal window of length: $s = 168$ hours, corresponding to weekly seasonality, the input sequence is represented as:

$$\mathbf{X}_t = [\mathbf{x}_{t-s+1}, \ldots, \mathbf{x}_t] \in \mathbb{R}^{s\times d}. \tag{1}$$

The forecasting objective consists in learning a parametric nonlinear mapping:

$$\hat{y}_{t+1} = f_\Gamma(\mathbf{X}_t), \tag{2}$$

where:

1) $\hat{y}_{t+1}$ is the predicted electricity consumption,
2) f_Γ is the deep forecasting model,
3) Γ denotes the trainable parameters.

The consumption series exhibits strong autocorrelation, regular weekly patterns, and relatively low variability. Such characteristics may limit the effectiveness of conventional statistical models and motivate the use of hybrid recurrent-attention architectures capable of modeling both local temporal dependencies and long-range interactions.

2.4. Hybrid LSTM–Multi-Head Attention Architecture

The proposed forecasting framework combines three complementary modules: LSTM encoder, Multi-Head Attention (MHA) and LSTM decoder. The global architecture is designed to jointly capture:

1) Short-term temporal dependencies,
2) Weekly seasonality,
3) Long-range nonlinear interactions,
4) Structured recurrent patterns.

2.4.1. LSTM Encoder

The encoder transforms the multivariate input sequence into a latent temporal representation:

$$\mathbf{h}_\tau = \mathrm{LSTM}(\mathbf{x}_\tau, \mathbf{h}_{\tau-1}), \quad \tau = t-s+1, \dots, t. \tag{3}$$

Each hidden state $\mathbf{h}_\tau \in \mathbb{R}^m$, contains temporal information accumulated from previous observations. The complete latent representation matrix is defined as:

$$H = \begin{bmatrix} h_{t-s+1} \\ h_{t-s+2} \\ \vdots \\ h_t \end{bmatrix} \in \mathbb{R}^{s\times m}. \tag{4}$$

This matrix contains the temporal embeddings extracted from the historical sequence and serves as input to the attention mechanism.

2.4.2. Multi-Head Attention Block

The Multi-Head Attention mechanism enables the model to identify the most informative temporal positions for forecasting future electricity consumption. It is particularly useful for capturing weekly periodicity and long-range dependencies [17].

For each attention head:

$$\mathrm{head}_i(Q^i, K^i, V^i) = \mathrm{softmax}\left(\frac{Q^i (K^i)^\top}{\sqrt{d_k}}\right) V^i, \tag{5}$$

where:

$$Q^i = HW_Q^i, \quad K^i = HW_K^i, \quad V^i = HW_V^i.$$

The outputs of all attention heads are concatenated:

$$Z = \mathrm{Concat}(\mathrm{head}_1, \dots, \mathrm{head}_M) W^O, \tag{6}$$

where:

1) M is the number of attention heads,
2) W^O is the projection matrix.

The attention block allows the model to dynamically focus on relevant temporal patterns and recurrent consumption behaviors.

2.4.3. LSTM Decoder and Output Layer

The decoder receives the attention-enhanced representation and reconstructs a compact temporal dynamic suitable for prediction:

$$\mathbf{z}_t = \mathrm{LSTM}_{dec}(Z). \tag{7}$$

The final prediction is obtained through a dense layer:

$$\hat{y}_{t+1} = W_o \mathbf{z}_t + b. \tag{8}$$

2.5. Training Procedure and Hyperparameter Optimization

The model parameters are optimized by minimizing the Mean Squared Error (MSE):

$$\mathcal{L}(\Gamma) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{9}$$

The training process uses the Adam optimizer with adaptive learning rate scheduling and early stopping.

The following hyperparameters are optimized: Encoder LSTM units, Decoder LSTM units, Number of attention heads, Attention key dimension, Dropout rate, Learning rate and Batch size.

A distributed hyperparameter search strategy is executed on HPC infrastructure using multiprocessing across multiple computing nodes.

2.6. Experimental Evaluation

The dataset is divided chronologically into: training set: 2014–2017, validation set: January–September 2018 and future forecasting period: October–December 2018.

Several evaluation metrics are considered: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric MAPE (sMAPE), Coefficient of Variation of RMSE (CVRMSE), Normalized RMSE (NRMSE) and Coefficient of Determination (R^2). In addition to forecasting accuracy, the study includes: statistical significance analysis using the Wilcoxon signed-rank test, ablation study, robustness analysis under noisy and missing data conditions.

3. Results

The experimental results for short-term electricity consumption forecasting are obtained using a weekly time step of s = 168 hours. The experimental protocol includes five independent random seeds, an ablation study, robustness tests under noisy and missing-data conditions, and statistical comparison using the Wilcoxon signed-rank test. All experiments were conducted with the target variable `CONSOMMATION_TOTALE`. The main objective of this study is to determine whether incorporating meteorological and demographic data degrades the performance of an LSTM model, or whether a hybrid architecture such as one based on attention mechanisms offers a significant improvement over simpler recurrent configurations.

3.1. Experimental Configuration

All models were trained using a weekly input window $s = 168$ hours, which corresponds to one full week of hourly observations. This choice allows the models to exploit daily and weekly consumption patterns. The training was conducted over the period 2014–2017, while the year 2018 was used for validation and testing. Five independent runs were performed using the seeds 11, 22, 33, 44, and 55. The optimized hyperparameters used in the experiments were: encoder units = 32, decoder units = 96, number of attention heads = 2, key dimension = 16, dropout = 0.3, learning rate = 0.001, batch size = 128, and 60 training epochs. Four model configurations were evaluated:

1) *LSTM-only*: recurrent model without attention;
2) *LSTM-decoder*: recurrent encoder followed by a decoder LSTM;
3) *LSTM-attention*: recurrent encoder followed by a Multi-Head Attention block;
4) *Full hybrid*: complete LSTM–Multi-Head Attention–LSTM architecture.

3.2. Main Forecasting Results

Table 2 reports the validation results averaged over five independent runs under clean-data conditions. The results show that all configurations achieve very high forecasting accuracy, with R^2 values greater than 0.9998 and MAPE values below 0.004%.

Table 2. *Validation performance under clean-data conditions averaged over five seeds.*

Model	MAE	RMSE	MAPE (%)	R^2
Full hybrid	253.53	313.63	0.00304	0.999908
LSTM-attention	259.24	312.34	0.00311	0.999909
LSTM-decoder	268.34	339.31	0.00321	0.999892
LSTM-only	290.75	365.31	0.00348	0.999873

The full hybrid model provides the lowest MAE and MAPE, while the LSTM-attention configuration gives the lowest RMSE by a very small margin. These results indicate that attention-based architectures slightly improve the forecasting accuracy when evaluated over multiple independent runs. However, the differences remain small in absolute terms, confirming that the electricity consumption series is highly structured and strongly predictable from recent temporal history. Figure 3 presents the validation MAPE comparison.

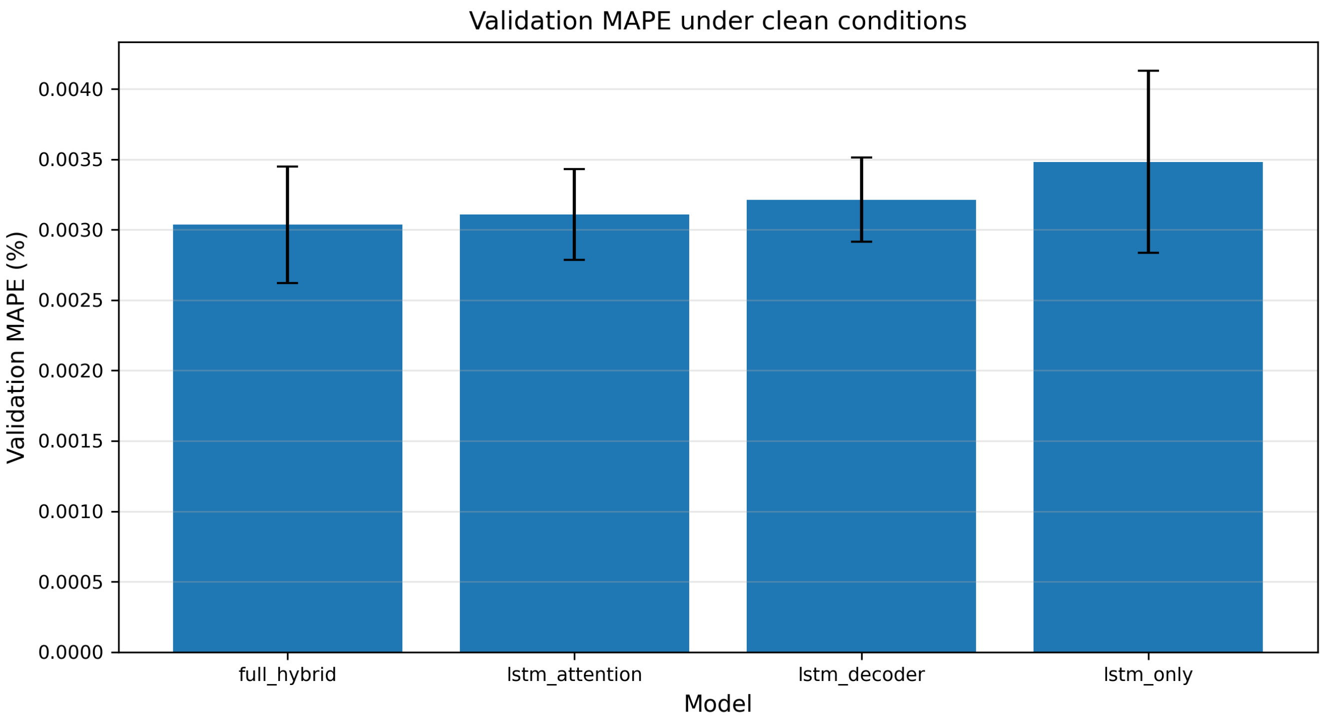

Figure 3. *Validation MAPE comparison under clean-data conditions.*

All models achieve extremely low percentage forecasting errors, confirming the highly structured nature of the electricity consumption series. Figure 4 presents the coefficient of determination obtained by each architecture.

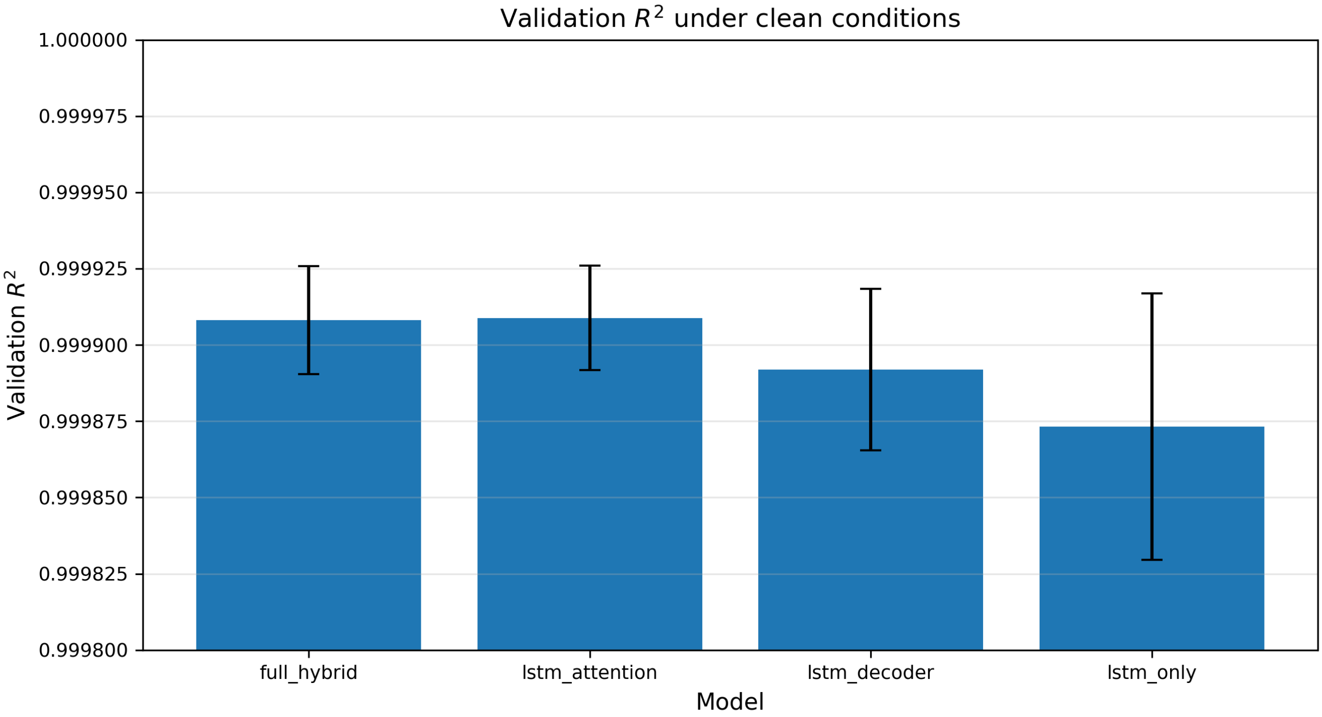

Figure 4. *Validation R^2 comparison under clean-data conditions.*

All architectures achieve: $R^2 > 0.9998$, indicating near-perfect reconstruction of the temporal dynamics.

3.3. Ablation Study

The ablation study quantifies the contribution of each architectural component. Table 3 compares the clean-condition validation performance of the four configurations.

Table 3. *Ablation study of recurrent, attention, and decoder components.*

Configuration	Role tested	RMSE	MAPE (%)	CVRMSE (%)
LSTM-only	Recurrent memory only	365.31	0.00348	0.00438
LSTM-decoder	Effect of decoder	339.31	0.00321	0.00407
LSTM-attention	Effect of attention	312.34	0.00311	0.00374
Full hybrid	Attention + decoder	313.63	0.00304	0.00376

The ablation results show that adding a decoder improves the LSTM-only baseline, while adding attention provides a larger gain in terms of RMSE and CVRMSE. The full hybrid architecture obtains the best MAE and MAPE, but its RMSE is almost identical to that of the LSTM-attention model. This suggests that the attention mechanism contributes more strongly to performance improvement than the final decoder in this specific setting. Figure 5 compares the RMSE obtained by the different configurations.

The results show that adding the decoder improves the standalone LSTM baseline, while the attention mechanism produces the largest performance gain. Figure 6 presents the MAPE comparison for the ablation configurations.

The full hybrid architecture achieves the lowest MAPE, indicating better relative forecasting precision.

3.4. Statistical Comparison

To assess whether the observed differences are statistically significant, a Wilcoxon signed-rank test was applied to the paired validation errors obtained across the five independent seeds. Table 4 reports the main pairwise comparisons using RMSE and MAPE.

Table 4. *Wilcoxon signed-rank test for paired model comparison under clean conditions.*

Model A	Model B	Metric	p-value	Better mean performance
Full hybrid	LSTM-only	RMSE	0.1250	Full hybrid
Full hybrid	LSTM-only	MAPE	0.3125	Full hybrid
Full hybrid	LSTM-attention	RMSE	1.0000	LSTM-attention
Full hybrid	LSTM-attention	MAPE	0.6250	Full hybrid
LSTM-attention	LSTM-only	RMSE	0.1250	LSTM-attention
LSTM-attention	LSTM-only	MAPE	0.3125	LSTM-attention
LSTM-decoder	LSTM-only	RMSE	0.4375	LSTM-decoder
LSTM-decoder	LSTM-only	MAPE	0.6250	LSTM-decoder

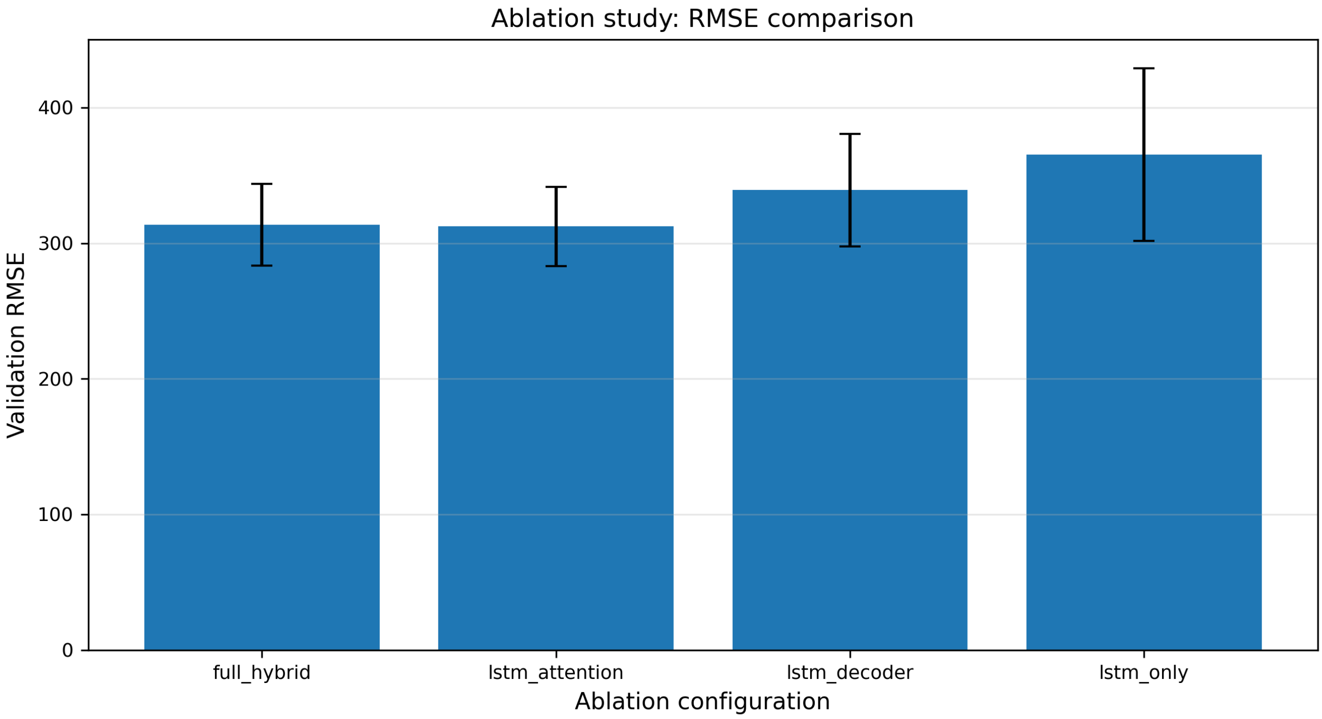

Figure 5. *Ablation study: RMSE comparison.*

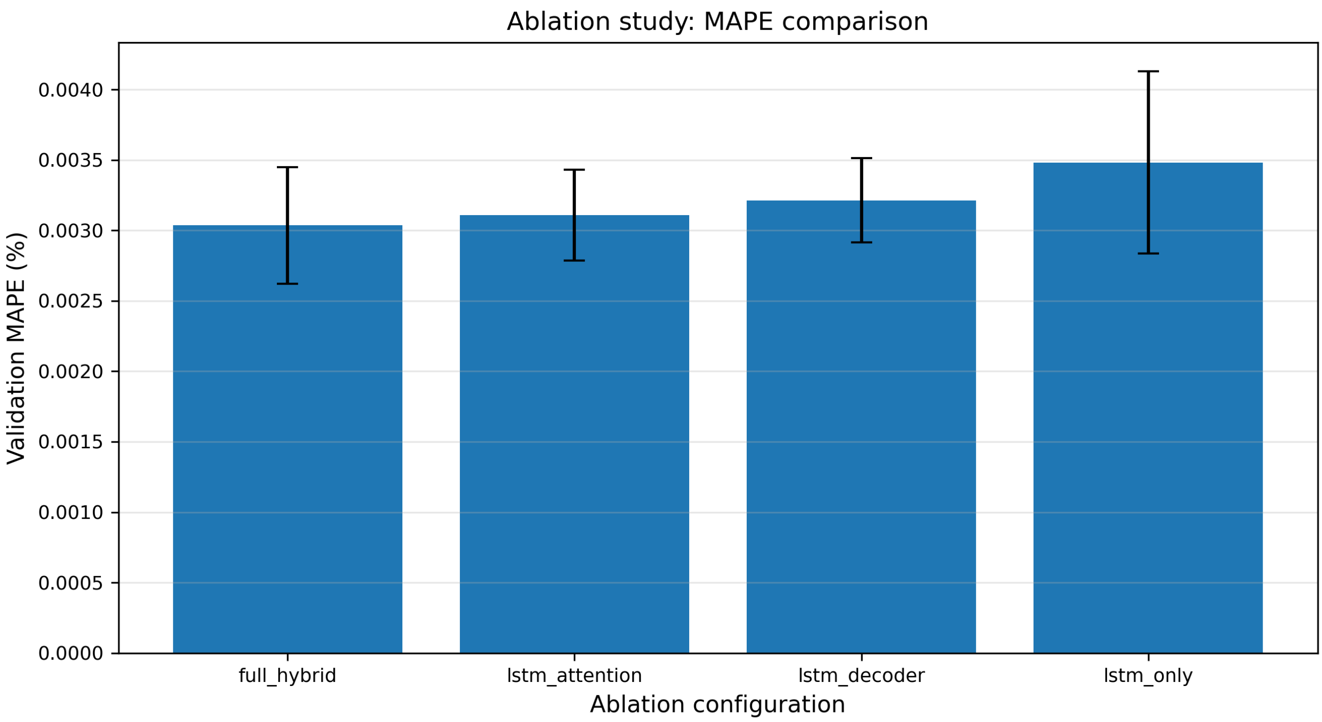

Figure 6. *Ablation study: MAPE comparison.*

Although the attention-based and hybrid models show better mean performance than the LSTM-only model, the Wilcoxon tests do not indicate statistically significant differences at the 5% level. This is mainly due to the small number of paired runs ($n = 5$) and the very low variability of the forecasting task. Therefore, the observed performance improvements should be interpreted as consistent but not statistically conclusive. Figure 7 presents the Wilcoxon p-values obtained for RMSE comparison.

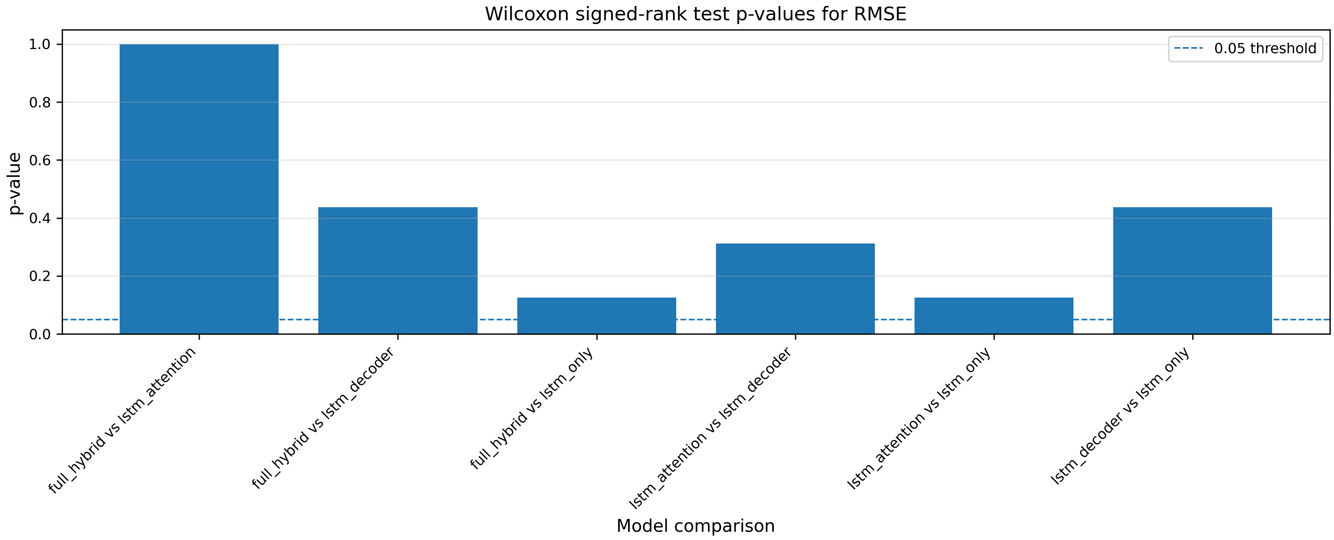

Figure 7. *Wilcoxon signed-rank test p-values for RMSE comparison.*

The results indicate that although the hybrid architectures achieve lower average errors, the performance differences are not statistically significant at the 5% significance level. Figure 8 presents the statistical comparison based on MAPE.

The absence of statistical significance is mainly explained by:

1) the very low forecasting variability,
2) the highly structured nature of the time series,
3) the small number of paired runs.

3.5. Robustness to Noisy Inputs

A robustness analysis was conducted by injecting Gaussian noise into the input features at levels of 5%, 10%, and 20%. Table 5 summarizes the RMSE and MAPE variations under noisy conditions.

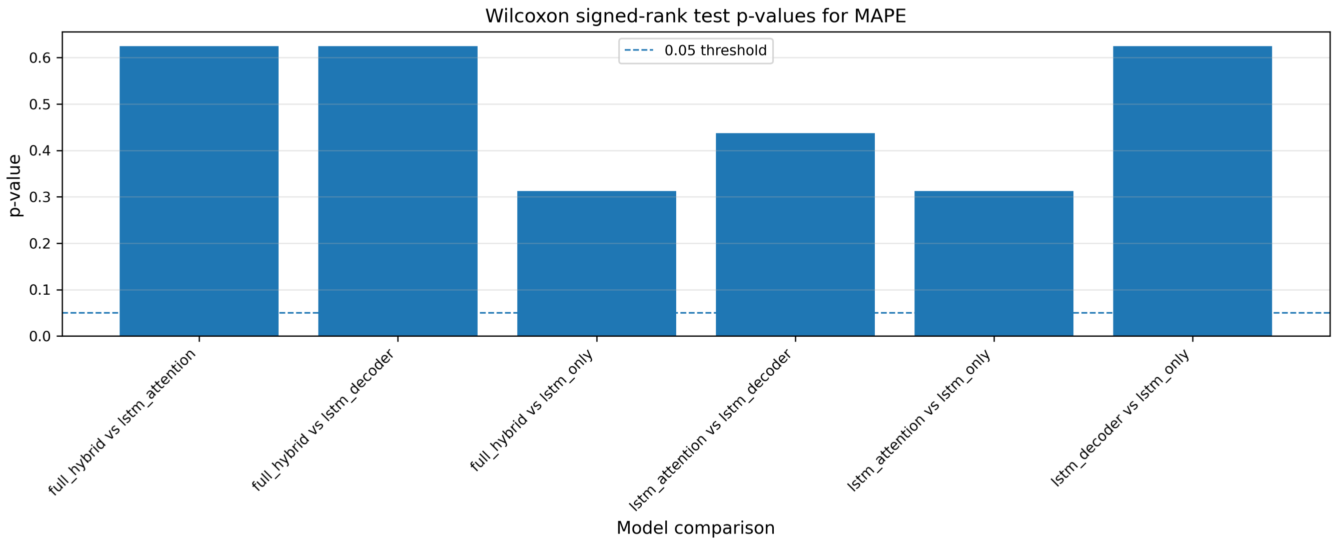

Figure 8. *Wilcoxon signed-rank test p-values for MAPE comparison.*

Table 5. *Robustness to noisy input features. Values are averaged over five seeds.*

Model	Noise level	RMSE	MAPE (%)
Full hybrid	0%	313.63	0.00304
Full hybrid	5%	14158.95	0.13562
Full hybrid	10%	28308.86	0.27117
Full hybrid	20%	56614.18	0.54228
LSTM-only	0%	365.31	0.00348
LSTM-only	5%	14158.21	0.13560
LSTM-only	10%	28308.89	0.27116
LSTM-only	20%	56615.36	0.54230

The results show that noise strongly degrades the forecasting performance of all models. The degradation is approximately proportional to the noise level, indicating that the models are highly sensitive to perturbations in the explanatory variables. Under noisy conditions, the performance differences between architectures become negligible, suggesting that input quality dominates architectural effects. Figure 9 presents the RMSE evolution as a function of the noise level.

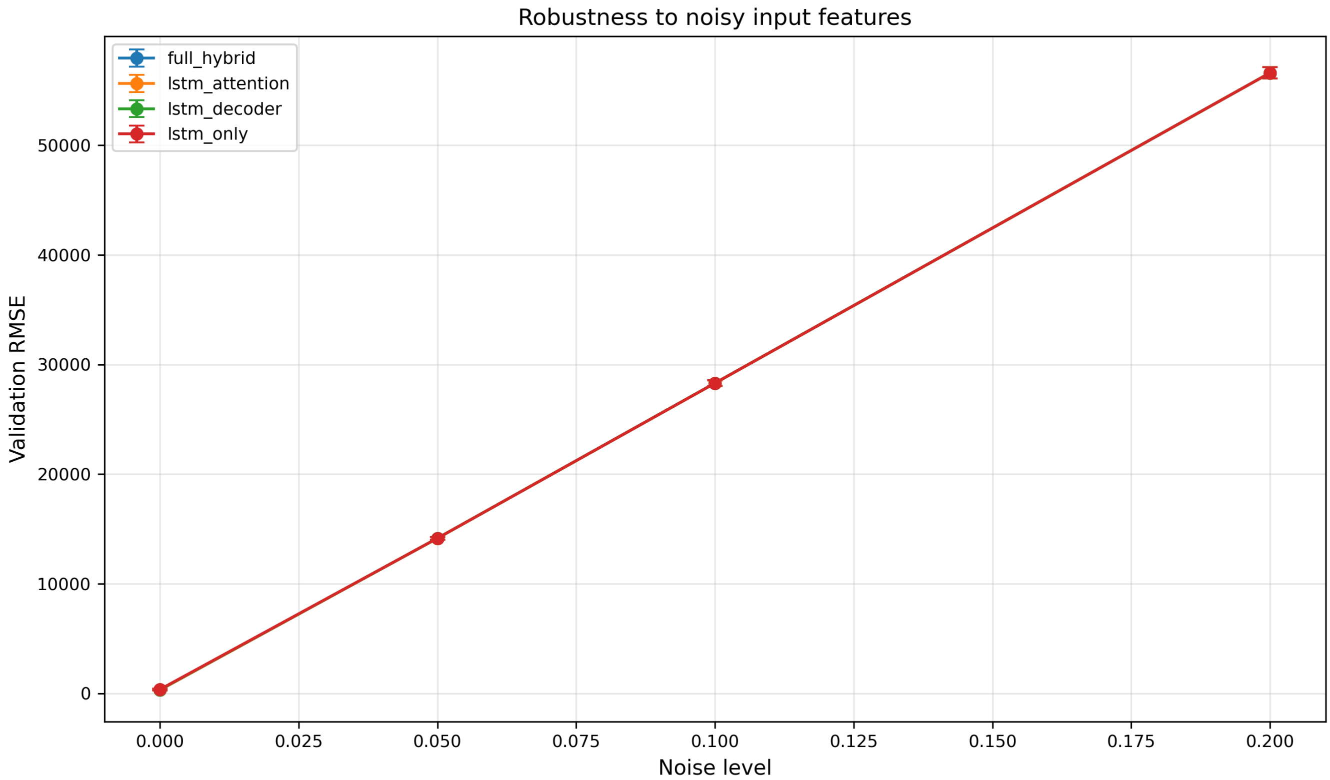

Figure 9. *RMSE evolution under noisy input conditions.*

The forecasting errors increase almost linearly with the noise level, showing that data quality strongly affects forecasting accuracy. Figure 10 presents the MAPE evolution under noisy conditions.

Under noisy conditions, all architectures exhibit very similar degradation behavior, indicating that input perturbations dominate architectural effects.

3.6. Robustness Analysis Under Missing Data Conditions

A second robustness analysis was performed by randomly removing 5%, 10%, and 20% of the input values and reconstructing them through interpolation. Table 6 reports the corresponding validation performance.

Table 6. *Robustness to missing input data. Values are averaged over five seeds.*

Model	Missing rate	RMSE	MAPE (%)
Full hybrid	0%	313.63	0.00304
Full hybrid	5%	340.93	0.00333
Full hybrid	10%	311.19	0.00306
Full hybrid	20%	357.98	0.00345
LSTM-only	0%	365.31	0.00348
LSTM-only	5%	377.67	0.00374
LSTM-only	10%	339.72	0.00332
LSTM-only	20%	334.76	0.00329

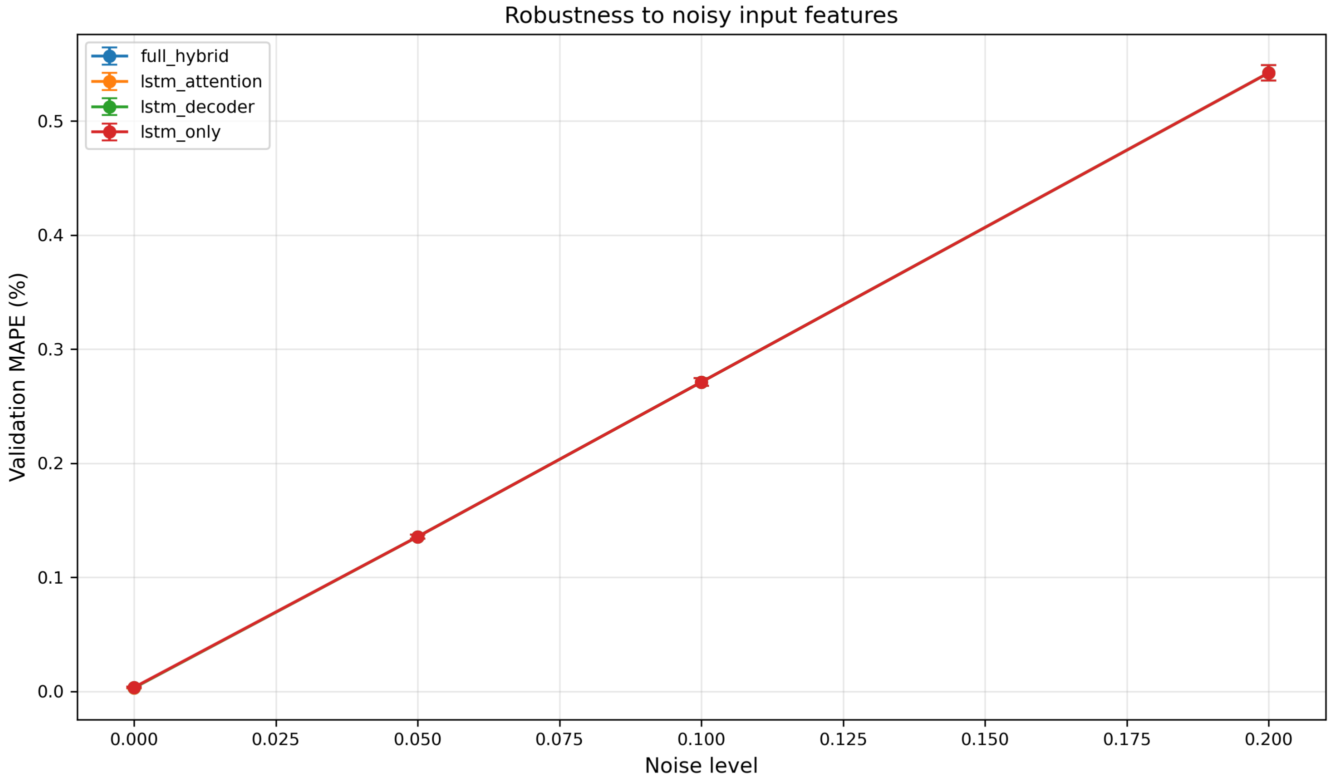

Figure 10. *MAPE evolution under noisy input conditions.*

Unlike input noise, missing values followed by interpolation produce only limited degradation. This indicates that the forecasting pipeline is relatively robust to moderate missingness when temporal interpolation is applied. The full hybrid model remains competitive under missing-data scenarios, while the LSTM-only model also maintains strong predictive accuracy. Figure 11 presents the RMSE evolution under missing-data conditions.

The degradation produced by missing data remains moderate compared with noisy conditions. Figure 12 presents the corresponding MAPE evolution.

The results indicate that interpolation-based reconstruction preserves most temporal structures required for forecasting.

3.7. Energy-oriented Interpretation

From an energy management perspective, lower RMSE and CVRMSE values imply reduced forecasting uncertainty, better short-term production scheduling, and improved load-following planning. Under clean conditions, the full hybrid and LSTM-attention models achieve the lowest error levels, while all configurations maintain extremely low relative errors. This suggests that the proposed forecasting pipeline can support operational decision-making for electricity demand management in Togo, especially in contexts where hourly forecasts are required for planning, reserve allocation, and demand-supply balancing.

Figure 13 presents the operational uncertainty comparison using CVRMSE.

The attention-based architectures produce slightly lower operational uncertainty than standalone recurrent models.

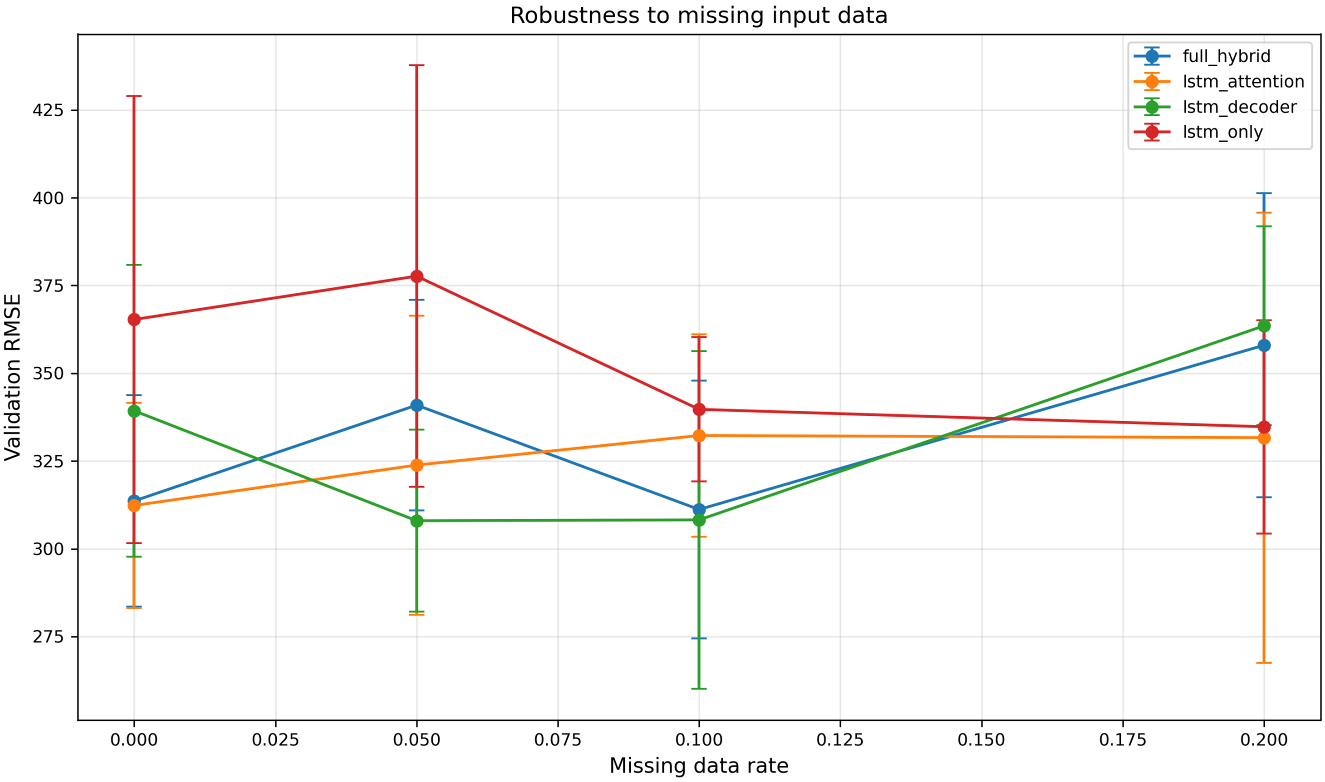

Figure 11. *RMSE evolution under missing-data conditions.*

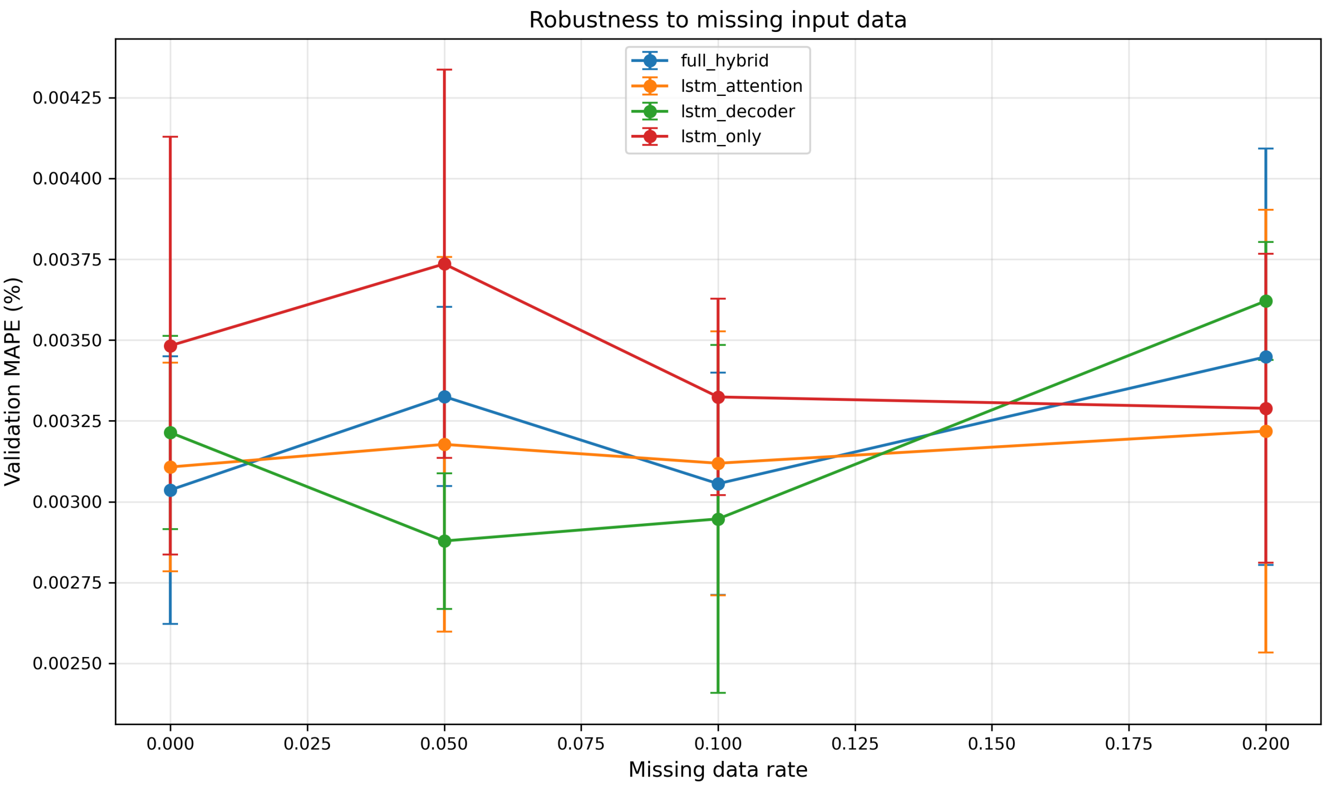

Figure 12. *MAPE evolution under missing-data conditions.*

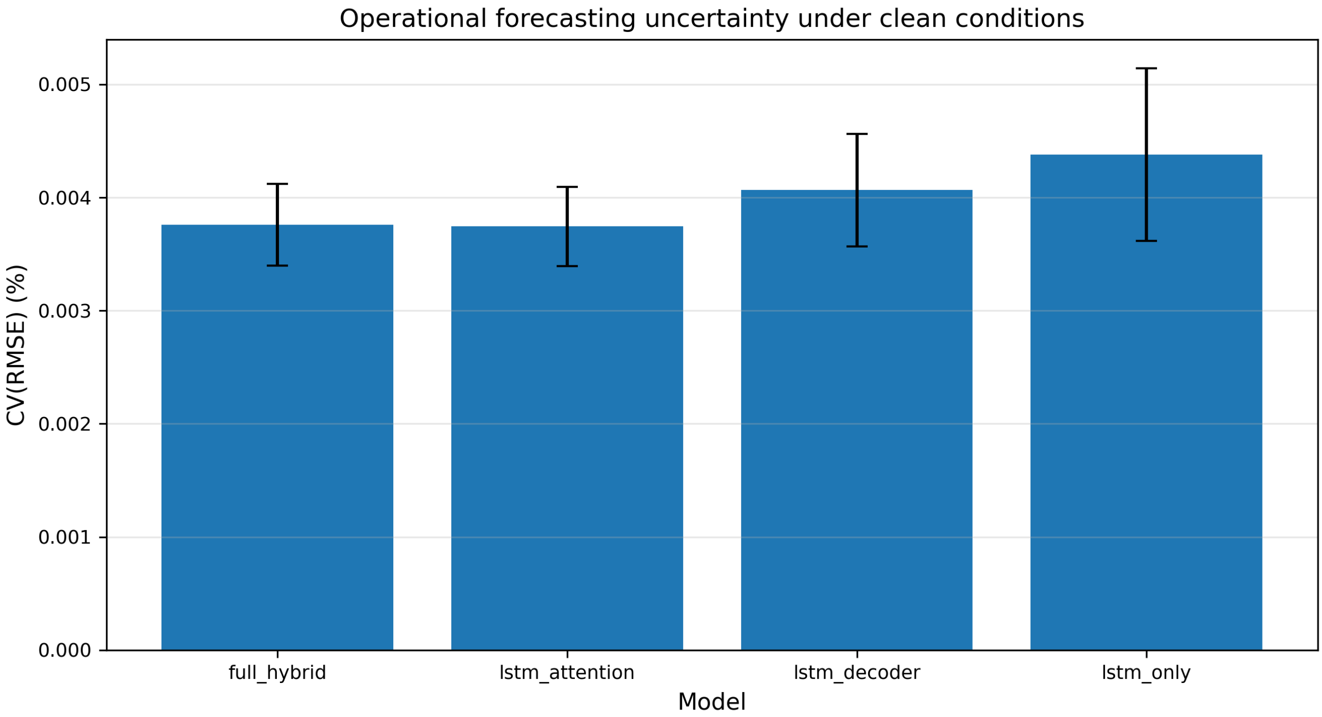

Figure 13. *Operational forecasting uncertainty measured using CVRMSE.*

3.8. Hourly Forecasting Analysis

Figures 14 and 15 compare observed and predicted hourly electricity consumption during the validation period for the LSTM-only and hybrid architectures, respectively.

Both figures show a strong agreement between observed and predicted electricity consumption. The predicted curves closely follow the real hourly dynamics, confirming the ability of both models to capture daily fluctuations and short-term consumption variations. The LSTM-only model presents slightly smoother predictions, whereas the hybrid model reproduces the same global dynamics with marginal local deviations. Overall, the visual analysis is consistent with the quantitative results, where both models achieve very low prediction errors.

3.9. Weekly Consumption Profile Analysis

Figures 16 and 17 present the weekly mean consumption profiles obtained for the LSTM-only and hybrid models.

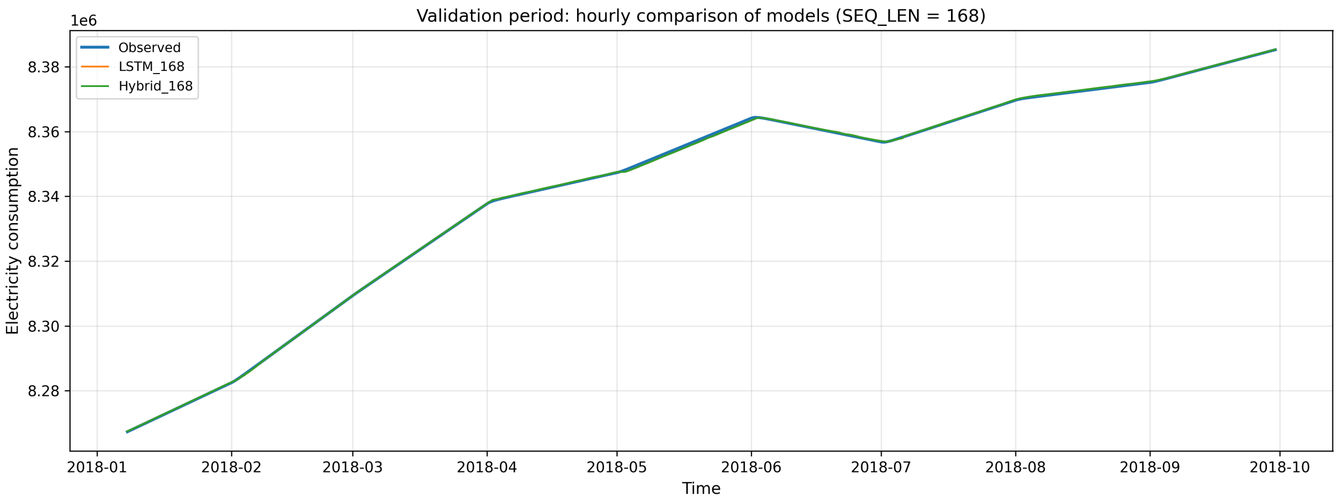

Figure 14. *Hourly observed and predicted electricity consumption during the validation period for the LSTM-only model.*

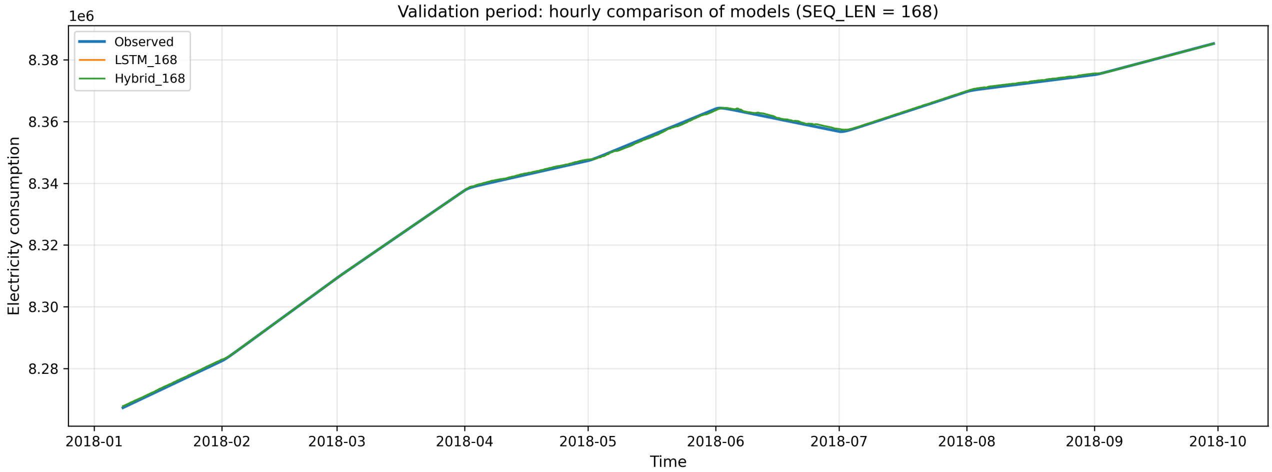

Figure 15. *Hourly observed and predicted electricity consumption during the validation period for the hybrid LSTM–MHA–LSTM model.*

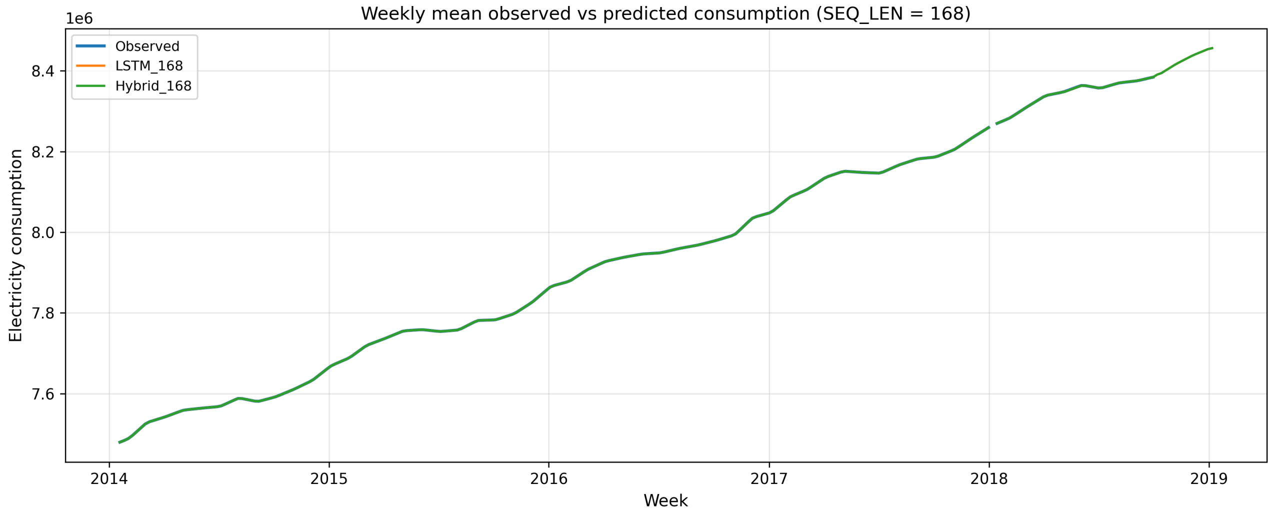

Figure 16. *Weekly mean observed and predicted electricity consumption profiles for the LSTM-only model.*

The weekly profiles demonstrate that both architectures preserve the medium-term temporal structure of electricity consumption. The predicted weekly averages remain close to the observed values, indicating that the models are able to reproduce not only local hourly variations but also broader weekly demand patterns. This confirms the relevance of using a weekly sequence length of 168 hours for short-term electricity consumption forecasting in Togo.

3.10. Summary of Findings

The results lead to the following observations:

1) All models achieve very high forecasting accuracy under clean conditions.
2) Attention-based configurations slightly improve average performance compared with LSTM-only.
3) The Wilcoxon test does not confirm statistically significant superiority at the 5% level.
4) Input noise severely degrades all models, indicating the importance of data quality.
5) Missing values reconstructed by interpolation have limited impact on performance.
6) The forecasting task remains strongly governed by temporal regularity and persistence.

4. Discussion

The experimental results demonstrate that all evaluated architectures achieve extremely high forecasting accuracy for short-term electricity consumption prediction in Togo. The very low forecasting errors obtained across all configurations confirm that the studied electricity demand series is highly structured, strongly autocorrelated, and governed by recurrent

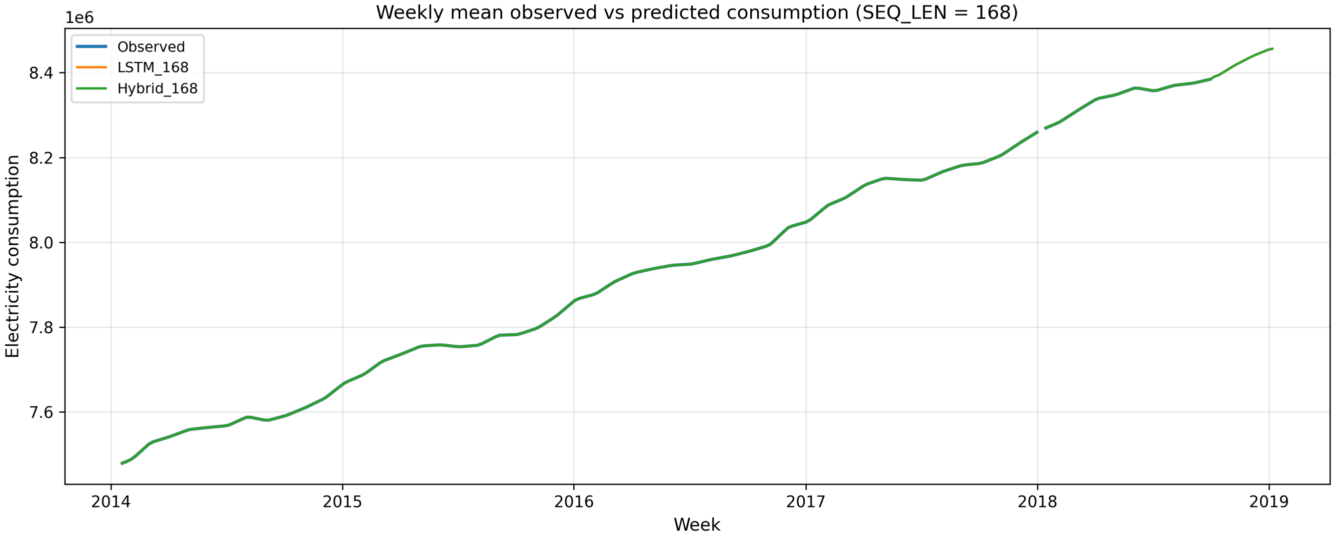

Figure 17. *Weekly mean observed and predicted electricity consumption profiles for the hybrid LSTM–MHA–LSTM model.*

temporal dynamics. Such characteristics are typical of electricity consumption profiles exhibiting strong daily and weekly periodicity.

The standalone recurrent architecture already provides excellent predictive performance, indicating that the recent temporal history contains most of the information required for accurate forecasting. This observation is supported by the very high coefficients of determination: $R^2 > 0.9998$, obtained for all models. Consequently, the forecasting task appears to be dominated by temporal persistence and recurrent seasonal structures.

Although all architectures produce highly accurate forecasts, the attention-based configurations slightly improve the average forecasting performance. In particular, the hybrid LSTM–MHA–LSTM architecture achieves the lowest MAE and MAPE values, while the LSTM-attention model produces the lowest RMSE. These improvements suggest that the attention mechanism helps identify relevant temporal positions and recurrent consumption patterns over the weekly input horizon.

The ablation study further clarifies the role of each architectural component. The addition of the decoder improves the baseline LSTM model, while the attention mechanism contributes the largest reduction in RMSE and CVRMSE. This indicates that attention mechanisms can enhance temporal feature selection and improve the representation of long-range dependencies. However, the performance gain remains relatively small compared with the already strong baseline performance of the standalone recurrent model.

The statistical comparison based on the Wilcoxon signed-rank test reveals that the observed improvements are not statistically significant at the 5% significance level. Several factors may explain this result. First, the forecasting task itself is characterized by very low variability due to the strong regularity of the electricity demand series. Second, the number of paired experimental runs remains limited. Third, the differences between architectures are extremely small in absolute magnitude. Therefore, although the hybrid architectures consistently produce lower mean errors, the superiority of attention-based models cannot be considered statistically conclusive in this specific context.

The robustness analysis highlights the critical importance of data quality for electricity consumption forecasting. The injection of Gaussian noise strongly degrades forecasting accuracy for all architectures, with forecasting errors increasing almost proportionally to the noise level. This behavior indicates that perturbations affecting meteorological or consumption variables directly impact the temporal representations learned by the models. Under noisy conditions, the differences between architectures become negligible, suggesting that data quality dominates architectural complexity.

Conversely, the models remain relatively robust to missing data when interpolation techniques are applied before training. Even under 20% missing-data conditions, the degradation remains moderate compared with the noise scenario. This result confirms that temporal interpolation preserves most of the structural information required for forecasting highly regular electricity demand series.

From a methodological perspective, the results suggest that complex hybrid attention-based architectures are not always necessary for highly structured electricity consumption forecasting tasks. Simpler recurrent architectures may provide a better trade-off between: forecasting accuracy, computational complexity, training stability and operational deployment.

This observation is particularly important for operational forecasting systems deployed in developing countries, where computational resources may remain limited. The weekly input horizon of: $SEQ\_LEN = 168$ hours, also plays a central role in the forecasting performance. The results confirm that weekly temporal memory is sufficient to capture the dominant electricity consumption dynamics observed in Togo. The models successfully reproduce: hourly fluctuations, daily consumption cycles, weekly periodicity and medium-term consumption trends.

The integration of meteorological and demographic information further improves the realism of the forecasting framework. Although electricity demand remains strongly driven by temporal persistence, exogenous variables contribute to capturing environmental and socioeconomic influences affecting consumption behavior. This multimodal integration is particularly relevant for African electricity systems, where weather conditions and demographic growth significantly influence energy demand evolution.

From an energy-management perspective, accurate short-term forecasting contributes directly to: production scheduling, reserve allocation, load balancing, operational planning, demand-response management. The very low CVRMSE values obtained in this study indicate reduced forecasting uncertainty, which may improve operational efficiency in the Togolese power system.

5. Conclusion

This study investigated short-term electricity consumption forecasting in Togo using recurrent and hybrid deep learning architectures combining LSTM networks and Multi-Head Attention mechanisms. The proposed framework integrates electricity consumption, meteorological, demographic, and temporal information in order to model the complex dynamics governing electricity demand.

The experimental results demonstrated that all evaluated architectures achieved extremely high forecasting accuracy, with very low RMSE and MAPE values and coefficients of determination exceeding 0.9998. The standalone LSTM model already provided excellent predictive performance, confirming that the studied electricity demand series is highly structured, strongly autocorrelated, and dominated by recurrent weekly patterns. The hybrid attention-based architectures slightly improved the average forecasting performance, particularly in terms of MAE and MAPE, although the statistical analysis did not confirm significant superiority at the 5% significance level.

The ablation study highlighted the contribution of the attention mechanism and decoder structure, while the robustness analysis emphasized the critical importance of data quality for electricity forecasting systems. The models remained relatively stable under missing-data conditions but were highly sensitive to noisy inputs.

Overall, the results demonstrate that recurrent deep learning architectures constitute effective and reliable tools for operational electricity demand forecasting in Togo and similar West African energy systems.

ORCID

0000-0002-1212-759X (Sena Apeke)
0000-0003-4023-6520 (Yao Bokovi)
0009-0001-3952-3373 (Kodjovi Senanou Gbafa)

Abbreviations

ANFIS	Adaptive Neuro-Fuzzy Inference System
ARIMA	AutoRegressive Integrated Moving Average
LSTM	Long Short-Term Memory
MHA	Multi-Head Attention
RNN	Recurrent Neural Network

Acknowledgments

The authors want to thank the Centre d'Excellence Régional pour la Maîtrise de l'Electricité (CERME) of University of Lome for providing support and enabling environment during this research.

Author Contributions

Yao Bokovi: Data curation, Investigation, Resources, Validation, Writing – review & editing
Kodjovi Senanou Gbafa: Formal Analysis, Validation, Writing – review & editing
Komlan Kolegain: Validation, Writing – review & editing
Sanoussi Ouro-Djobo: Supervision, Project administration, Validation, Writing – review & editing

Data Availability Statement

The electricity consumption data used in this study were provided by the Compagnie Énergie Électrique du Togo (CEET) and are not publicly available due to institutional and confidentiality restrictions. Meteorological data are publicly available from the Open-Meteo platform, and demographic data are available from the World Bank database. Access to the electricity consumption data may be granted upon reasonable request and subject to CEET authorization.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Dinata, S., Azka, M., Muliady, F., et al., Short-Term Load Forecasting Double Seasonal ARIMA Methods: An Evaluation Based on Mahakam-East Kalimantan Data, *AIP Conference Proceedings*, vol. 2268, article 020004, 2020, https://doi.org/10.1063/5.0017643

[2] Open-Meteo, Archive Weather API, *Open-Meteo Documentation*, 2024, Available: https://archive-api.open-meteo.com/ Accessed 16 May 2026

[3] World Bank, World Development Indicators, *World Bank Open Data*, 2024, Available: https://donnees.banquemondiale.org/ Accessed 16 May 2026

[4] Nguyen, Q., Nguyen, N., Tran Ngoc, T., et al., Online SARIMA Applied for Short-Term Electricity Load Forecasting, *Research Square*, pp. 1003–1019, 2023, https://doi.org/10.21203/rs.3.rs-439120/v1

[5] Ajlouni, S. A., Using ARIMA Model to Forecast Electricity Load in Jordan, *Jordan Journal of Earth and Environmental Sciences*, vol. 11, no. 2, pp. 127–139, 2024, https://doi.org/10.35516/jjes.v11i2.1733

[6] Matos, M., Almeida, J., GonÃ§alves, P., et al., A Machine Learning-Based Electricity Consumption Forecast and Management System for Renewable Energy Communities, *Energies*, vol. 17, article 630, 2024, https://doi.org/10.3390/en17030630

[7] Fan, S. and Hyndman, R. J., Short-Term Load Forecasting Based on a Semi-Parametric Additive Model, *IEEE Transactions on Power Systems*, vol. 27, pp. 134–141, 2012, https://doi.org/10.1109/TPWRS.2011.2162082

[8] Hong, T., Pinson, P., and Fan, S., Global Energy Forecasting Competition 2012, *International Journal of Forecasting*, vol. 30, pp. 357–363, 2014, https://doi.org/10.1016/j.ijforecast.2013.07.001

[9] Various Authors, Long Short-Term Memory Networks: A Comprehensive Survey, *Engineering Proceedings*, vol. 6, article 215, 2025, https://doi.org/10.3390/engproc202500609215

[10] Hong, T. and Fan, S., Probabilistic Energy Forecasting, *International Journal of Forecasting*, vol. 32, pp. 896–913, 2016, https://doi.org/10.1016/j.ijforecast.2016.02.001

[11] Weron, R., Electricity Price Forecasting: A Review, *International Journal of Forecasting*, vol. 30, pp. 1030–1081, 2014, https://doi.org/10.1016/j.ijforecast.2014.08.008

[12] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M., Time Series Analysis: Forecasting and Control, *John Wiley & Sons*, 5th ed., 712 pages, 2015

[13] Hochreiter, S. and Schmidhuber, J., Long Short-Term Memory, *Neural Computation*, vol. 9, pp. 1735–1780, 1997, https://doi.org/10.1162/neco.1997.9.8.1735

[14] Bokovi, Y., MoyÃ¨me Kabe, S., Sedzro Kwami, S., Takouda, P., and Lare, Y., Machine Learning Electrical Load Forecasting: An Application in Microgrid Energy Consumption with Adaboost Regressor Approach and a Comparative Study with Hybrid Method Based on LSTM and MLP Approaches, *Journal of Sustainable Development of Energy, Water and Environment Systems*, vol. 13, article 1130606, 2025, https://doi.org/10.13044/j.sdewes.d13.0606

[15] Suganthi, L. and Samuel, A. A., Energy Models for Demand Forecasting—A Review, *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 1223–1240, 2012, https://doi.org/10.1016/j.rser.2011.08.014

[16] Kong, W., Dong, Z. Y., Jia, Y., et al., Short-Term Residential Load Forecasting Based on LSTM, *IEEE Transactions on Smart Grid*, vol. 10, pp. 841–851, 2019, https://doi.org/10.1109/TSG.2017.2753802

[17] Vaswani, A., Shazeer, N., Parmar, N., et al., Attention Is All You Need, *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, https://doi.org/10.5555/3295222.3295349

[18] Ozcan, A., Catal, C., and Kasif, A., Energy Load Forecasting Using Attention RNN, *Sensors*, vol. 21, article 7115, 2021, https://doi.org/10.3390/s21217115

[19] Ding, S., He, D., and Liu, G., Improving Short-Term Load Forecasting with CNN and Multi-Head Attention, *Electronics*, vol. 13, article 5023, 2024, https://doi.org/10.3390/electronics13245023

[20] Hua, Q., Fan, Z., Mu, W., et al., CNN-GRU with Attention Mechanism, *Energies*, vol. 18, article 106, 2025, https://doi.org/10.3390/en18010106

[21] Wu, H., Xu, J., Wang, J., and Long, M., Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting, *arXiv preprint*, 2021, Available: https://doi.org/10.48550/arXiv.2106.13008

[22] Semekonawo, K. P. and Kam, S., Electricity Demand Forecasting in West Africa, *Journal of Energy Research and Reviews*, vol. 12, pp. 26–36, 2022, https://doi.org/10.9734/JENRR/2022/v12i230296

[23] Palanga, E. T. G., Bara, K. K. A., and Barate, M., Modeling of Electric Energy Consumption in Humid Zones Using Selected Meteorological Variables Through XGBoost, ANFIS, and RNN Approaches, *American Journal of Energy Engineering*, vol. 13, pp. 189–212, 2025, https://doi.org/10.11648/j.ajee.20251304.14

[24] Liu, J., Short-Term Load Forecasting of Electric Power System Based on Meteorological Factors, *Proceedings of ICMMBE 2016*, 2016, https://doi.org/10.2991/icmmbe-16.2016.38

[25] Cheng, Y., et al., Short-Term Load Forecasting Considering Improved Cumulative Effect of Hourly Temperature, *Electric Power Systems Research*, 2021, https://doi.org/10.1016/S0378-7796(21)00727-6

[26] Andoh, P. Y., Sekyere, C. K., et al., Forecasting Electricity Demand in Ghana, *Journal of Applied Engineering and Technological Science*, vol. 3, 2021, https://doi.org/10.37385/jaets.v3i1.288

[27] Adjamagbo, C., Salami, A., Bokovi, Y., et al., Forecasting of Electric Energy Consumption in Togo, *International Journal of Advanced Research*, vol. 8, pp. 22–28, 2020, https://doi.org/10.21474/IJAR01/11818

[28] Amega, K., Moumouni, Y., and Lare, Y., Electricity Consumption in Lomé, *International Journal of Energy and Power Engineering*, vol. 10, pp. 141–150, 2021, https://doi.org/10.11648/j.ijepe.20211006.17