Real-World Load and Wind Power Forecasting in Greek Island Energy Systems: A Deep Learning Perspective

Dimitrios-Christos Asimopoulos*†, Vasileios Vitsas†, Panagiotis Radoglou-Grammatikis‡§, Maria Fotopoulou¶, Nikolaos Andriopoulos¶, Georgios Lampsidis Tompros‖,and Panagiotis Sarigiannidis‡

***Abstract*—Accurate short-term forecasting of electricity load and renewable generation is essential for the reliable operation of energy systems. This paper presents a comparative evaluation of classical statistical methods and deep learning models for minute-level load and wind power forecasting using real-world data from the Greek islands of Samos and Ikaria. Autoregressive models (ARIMA, SARIMAX, AutoReg) are benchmarked against Long Short-Term Memory (LSTM) networks and Deep Neural Networks (DNNs). The models are evaluated using both standard one-step-ahead prediction and realistic rolling and recursive multi-step forecasting scenarios. Experimental results show that deep learning approaches significantly outperform statistical methods across all tasks, achieving up to one to two orders of magnitude lower error. Among them, the DNN demonstrates the most stable and robust behavior, especially under long-horizon deployment. The study also highlights the increased difficulty of wind power forecasting compared to load prediction due to strong intermittency and non-stationarity.**

***Index Terms*—Load Forecasting, Wind Power Forecasting, Deep Learning, Time Series Analysis, Island Energy Systems**

I. Introduction

The rapid penetration of renewable energy sources (RES) and the increasing electrification of modern societies have made accurate short-term forecasting a cornerstone of power system operation and planning. Reliable predictions of electricity demand and renewable generation are essential for ensuring grid stability, optimizing dispatch strategies, and improving the efficiency of energy management systems. This need becomes even more critical in island and weakly interconnected power systems, where limited reserves, reduced inertia, and strong dependence on local generation amplify the impact of forecasting errors.

Island energy systems represent particularly challenging environments for forecasting tasks. Their load profiles often exhibit strong seasonality and abrupt changes due to tourism and local activities, while renewable generation—especially wind power—is characterized by high intermittency and strong non-stationarity. In such systems, even small prediction errors can lead to inefficient dispatch decisions, increased reserve requirements, or threats to system reliability.

Traditionally, short-term load and generation forecasting has relied on statistical time series models such as Autoregressive Integrated Moving Average (ARIMA), seasonal ARIMA(SARIMAX), and AutoRegressive (AutoReg). Although these methods are computationally efficient and mathematically interpretable, their performance depends on assumptions of linearity and stationarity, which are increasingly violated in modern renewable-rich energy systems. As a result, their effectiveness is often limited in highly dynamic environments. Recent reviews confirm a clear shift toward deep learning-based short-term load forecasting in such renewable-rich, highly dynamic power systems [1].

In recent years, deep learning techniques have emerged as powerful alternatives for time series forecasting in complex and nonlinear settings. Models such as Long Short-Term Memory (LSTM) networks [2] and Deep Neural Networks (DNNs) have demonstrated strong ability to capture long-range dependencies and nonlinear dynamics, making them particularly suitable for modeling the volatile behavior of both electricity demand and renewable generation.

Despite growing interest in deep learning-based energy forecasting, several gaps remain. Many studies focus on large interconnected grids, where aggregation effects reduce volatility. Moreover, most evaluations are limited to one-step-ahead prediction, which does not reflect real operational deployment, where rolling or recursive multi-step forecasting is typically

This research was funded by the European Union's Horizon Europe Innovation Action, SINNOGENES (STORAGE INNOVATIONS FOR GREEN ENERGY SYSTEMS), under grant agreement No. 101096992. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

*Dimitrios Christos Asimopoulos is with MetaMind Innovations, Kila, 50100 Kozani, Greece — E-Mail: dasimopoulos@metamind.gr

†Dimitrios-Christos Asimopoulos and V.Vitsas are with the Department of Information and Electronic Engineering, International Hellenic University, Sindos Campus 57400, Thessaloniki, Greece – E-Mail: {dimiasim3, vitsas}@ihu.gr

‡P. Radoglou-Grammatikis and P. Sarigiannidis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, Campus ZEP Kozani, 50100 Kozani, Greece — E-Mail: {pradoglou, psarigiannidis}@uowm.gr

§P. Radoglou-Grammatikis is also with K3Y Ltd, Studentski District, Vitosha Quarter, Bl. 9, 1700 Sofia, Bulgaria — E-Mail: pradoglou@k3y.bg

¶M. Fotopoulou and N. Andriopoulos are with the Department of Research, Technology Development Independent Power Transmission Operator 89, Dyrrachiou str, Athens, 10443 — E-Mail: {m.fotopoulou, k.andriopoulos}@admie.gr

‖ G. Lampsidis Tompros is with the Department of Research and Innovation/Department of Non-Interconnected Islands, Hellenic Electricity Distribution Network Operator (H.E.D.N.O.), Perrevou 20 Kallirois 5, Athens, Greece - E-Mail: G.Lampsidis@deddie.gr

required and error accumulation becomes critical. Finally, systematic comparisons using high-resolution real-world data from challenging environments such as island grids remain limited.

This paper addresses these gaps through a comprehensive experimental evaluation of short-term load and wind power forecasting using minute-level real-world data from the Greek islands of Samos and Ikaria. Classical statistical methods (ARIMA, SARIMAX, AutoReg) are benchmarked against deep learning models (LSTM and DNN) under both standard one-step-ahead and realistic rolling and recursive forecasting scenarios.

The main contributions of this paper are summarized as follows:

- A systematic comparison between classical statistical methods and deep learning models for load and wind power forecasting using high-resolution real-world island data, sourced from the island's grid operator (Distribution System Operator / Network Operator).
- An extensive evaluation under both standard and realistic rolling and recursive multi-step forecasting scenarios, assessing long-horizon stability and error accumulation.
- A multi-dataset and multi-regime experimental analysis covering load demand and wind generation under grid and hybrid operation.
- An experimental demonstration that deep learning models, particularly a simple DNN architecture, consistently outperform statistical approaches and exhibit superior robustness.

The results show that deep learning models achieve substantially lower errors across all tasks, with the DNN exhibiting the most stable behavior under long-horizon deployment, while the LSTM remains highly competitive in one-step-ahead prediction. The analysis also confirms that wind power forecasting remains intrinsically more challenging than load forecasting due to strong intermittency and non-stationarity.

The remainder of this paper is organized as follows. Section II reviews the relevant literature. Section III presents the forecasting framework and models. Section IV describes the experimental setup and datasets. Section V reports the results, while Section VI discusses the findings. Finally, Section VII concludes the paper.

II. Relevant Literature

In literature there are several articles on advanced forecasting techniques for load but also for Wind Farm (WF) production, based on either statistical / classical approaches or on Machine Learning (ML). For example, Ullah et al. [3] have published a comprehensive review on short-term load forecasting, focusing on Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) approaches. Furthermore, Kumar Dubei et al. [4] have published a review paper on ML and hybrid intelligence for wind farm production forecasting, including Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Artificial Neural Networks (ANN), etc. Of note, WF production forecasting is considered to have a large error throughout literature because the production scales (roughly) with the cube of wind speed, which is quite unpredictable and fast-changing on its own. Specific contributions on both load and WF forecasting are analyzed below. On load forecasting, specifically on island-level, He et al. [5] performed microgrid load forecasting for future Electric Vehicle (EV) massive penetration in Okinawa island, Japan [6]. The authors used the Monte Carlo method for their scenarios, coupled with Genetic Algorithm (GA), considering the island's demand, demand-response schemes, the probability of parking time and its duration, etc. Andriopoulos et al. [7] propose a short-term load forecasting framework that combines statistical data transformation with deep learning techniques, emphasizing the role of preprocessing to enhance model performance. By exploiting time-series statistical properties and optimizing hyperparameters, the authors demonstrate that CNN-based models can achieve improved prediction accuracy compared to established LSTM approaches across multiple open datasets. Their results highlight the importance of feature engineering and data-driven model tuning as key enablers for accurate load forecasting in modern, RES-rich and highly variable power systems. In parallel, multi-scale ensemble deep learning frameworks have been proposed for short-term load forecasting, further improving accuracy over single deep models [8] Lopez et al. [9] studied the influence of meteorological variables on real-time Short-Term Load Forecasting (STLF) in the Balearic Islands, which is particularly useful for the local operator, including temperature, solar radiation, cloudiness and wind velocity. In [10] the authors focused on the development of a prediction model to be applied on ten main Spanish islands / systems: seven in the Canary Islands, and three in the Balearic Islands. The developed models, which constitute variations of ARIMA, are currently used by the Spanish TSO to make hourly one-day-ahead forecasts of the electricity demand. On WF production forecasting, the authors of [] propose an ML-based model for wind power ramp event forecasting. According to the study, Light Gradient Boosting Machine (LightGBM) combined with ensemble Numerical Weather Prediction (NWP) data improves ramp event prediction accuracy. The algorithm is applied on the Jeju island's WFs, located in South Korea. Moreover, Kumar et al. [4] studied the forecasting of wind speed over multiple horizons and found that: i) statistical models work well for very short forecasts but quickly lose accuracy over longer periods, ii) ML, especially XGBoost, offers more stable and reliable predictions, and iii) deep learning models deliver the best performance across short and long term forecasts. Moreover, the authors proved that by using Variational Mode Decomposition (VMD) with LSTM, they can further boost accuracy by filtering noise and capturing hidden wind patterns.

III. Methodology

This section outlines the methodological framework employed for short-term load and wind power forecasting using minute-level data. The approach is consistent with our previous

work on load forecasting [11], and it spans both statistical and deep learning paradigms. Specifically, we adopt and benchmark the following forecasting techniques: ARIMA, SARIMAX, AutoReg, LSTM networks, and DNNs. Additionally, a simple Moving Average model is used as a baseline sanity check.

A. Forecasting Framework

The forecasting task is formulated as a supervised time series problem, where a fixed-length sequence of past values is used to predict the next time step. Let $X_t = \{x_{t-L}, x_{t-L+1}, \ldots, x_{t-1}\}$ be the input and $y_t = x_t$ be the output. This sliding window approach is used for all models that require sequential input. Similar sliding-window formulations are widely adopted in recent deep learning and automated deep learning frameworks for load forecasting [12].

B. Autoregressive Integrated Moving Average (ARIMA)

ARIMA combines autoregression (AR), differencing (I), and moving average (MA). It is represented as:

$$Y_t = \mu + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{1}$$

where Y_t is the forecasted value, ϕ_i and θ_j are AR and MA parameters, and ε_t is the error term.

C. Seasonal ARIMA (SARIMAX)

SARIMAX extends ARIMA by including seasonal components. In this study, SARIMAX is used **without exogenous regressors** in order to ensure a fair comparison with purely data-driven models.

$$SARIMAX(p,d,q)(P,D,Q,s) \tag{2}$$

D. AutoRegressive (AutoReg)

The AutoReg model is a pure autoregressive model of order p:

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \tag{3}$$

It serves as a lightweight statistical baseline for comparison with more complex models.

E. Long Short-Term Memory (LSTM)

LSTM networks are a class of recurrent neural networks that maintain memory over long sequences via gate mechanisms:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{4}$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{5}$$
$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{6}$$
$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{7}$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{8}$$
$$h_t = o_t \cdot \tanh(C_t) \tag{9}$$

Our LSTM architecture consists of two stacked LSTM layers with 50 units, followed by a Dense output layer.

F. Deep Neural Network (DNN)

The DNN model is implemented as a fully connected feedforward neural network operating on flattened sliding-window input sequences. The architecture consists of two hidden layers with 64 and 32 neurons, respectively, using ReLU activation functions, followed by a single-neuron output layer with sigmoid activation to match the normalized target range.

Formally, given an input vector $\mathbf{x} \in \mathbb{R}^L$, the network computes:

$$\begin{aligned} \mathbf{h}_1 &= \mathrm{ReLU}(W_1\mathbf{x} + \mathbf{b}_1), \\ \mathbf{h}_2 &= \mathrm{ReLU}(W_2\mathbf{h}_1 + \mathbf{b}_2), \\ \hat{y} &= \sigma(W_3\mathbf{h}_2 + b_3). \end{aligned} \tag{10}$$

The model is trained using the Adam optimizer with mean squared error (MSE) as the loss function. Early stopping is employed based on validation loss to prevent overfitting. Despite its relatively simple structure, this architecture proved highly effective in capturing the short-term temporal dependencies and nonlinear dynamics of both load and wind time series.

G. Evaluation Metrics

We evaluate forecasting performance using three metrics:

- **Root Mean Squared Error (RMSE)**:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

- **Mean Absolute Error (MAE)**:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

- **Symmetric Mean Absolute Percentage Error (sMAPE)**:

$$\mathrm{sMAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|y_i| + |\hat{y}_i|)/2} \tag{13}$$

RMSE and MAE measure absolute prediction errors in normalized units, while sMAPE expresses the error as a percentage, providing a scale-independent measure that is more robust than the standard MAPE, especially when true values are close to zero. All models are evaluated on the same test data for consistency.

IV. Evaluation Analysis

A. Experimental Setup

As already described, this paper evaluates the performance of multiple forecasting techniques, including ARIMA, SARIMAX, LSTM, and DNNs, on minute-level wind power data. The experiments were implemented using the Python programming language. Data preprocessing tasks such as cleaning, normalization, and sequence generation were performed using the Pandas library, while model development was carried out using the `statsmodels` library for statistical models and

the Keras deep learning framework for LSTM and DNN implementations.

A look-back parameter of 60 was used to define the number of previous time steps provided as input to the forecasting models. For the deep learning models, a batch size of 64 and a maximum of 50 training epochs were initially configured. Early stopping was employed to terminate training based on convergence of the loss function. The models were trained using the Adam optimizer and mean squared error (MSE) as the loss function.

All experiments were conducted on a MacBook Air equipped with Apple's M2 chip, an 8-core CPU (4 performance and 4 efficiency cores), and 8 GB of unified memory. The device runs macOS Sequoia and does not include a discrete GPU, relying instead on the integrated Apple GPU for all computations. Despite its compact and energy-efficient architecture, the system provided sufficient performance for training and evaluating the forecasting models used in this study.

B. Dataset Description

For both the islands of Ikaria and Samos, hour granularity (i) load demand (MW) and (ii) wind production (MW) datasets were collected for the years 2023 and 2024 and then utilized for the training of the models and the comparisons of their output. It is important to note that depending on the electrical grid's operating conditions, the Greek Distribution System Operator (DSO) may curtail wind production and force Wind Parks to operate at lower capacities. As the focus of the models is to accurately predict actual wind production, the datasets described in (ii) were non-curtailed. For the curation and validation of the wind production data, wind speed data were also internally used as an accuracy index before being provided as input for the models. Load demand profiles were utilized in Sections A. and C. and wind production datasets in Sections B. and D.

C. Data Preprocessing

Prior to modeling, the following preprocessing steps were applied:

- **Negative value removal:** All negative power values were clipped to zero, as they are physically invalid in the context of wind power generation.
- **Normalization:** The target variable was normalized to the $[0, 1]$ range using MinMaxScaler, following the formula:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

- **Train-Test Split:** The dataset was split chronologically into 60% for training and 40% for testing to preserve temporal ordering and prevent data leakage.

D. Sequence Generation for Deep Learning Models

For LSTM (Figure 1) and DNN (Figure 2) models, input-output pairs were generated using a fixed-size sliding window. Each input sequence consists of the previous 60 minutes of wind power data, and the output is the value at the next minute. This results in a supervised learning setup with temporal context:

$$X_t = \{x_{t-60}, x_{t-59}, \ldots, x_{t-1}\}, \quad y_t = x_t$$

The sequences were reshaped appropriately for use with deep learning architectures: 3D for LSTM and flattened 2D for DNN.

E. Model Configurations

The forecasting models were configured with consistent design and tuning strategies to ensure fair comparison. Table **??** summarizes the architectural details and hyperparameters of each model used in the experiments. For statistical models, ARIMA was configured with an order of $(1, 1, 1)$, while SARIMAX followed a similar configuration without seasonal or exogenous components.

For deep learning models, the LSTM network comprised two stacked LSTM layers with 50 units each, followed by dropout regularization with a rate of 0.2, and a Dense output layer. The DNN model included two fully connected layers with 64 and 32 neurons respectively, using ReLU activations, and a final output layer with a sigmoid activation to match the normalized target range.

Both LSTM and DNN models were trained using the Adam optimizer with a learning rate of 0.001 and mean squared error (MSE) as the loss function. A batch size of 64 and a maximum of 50 training epochs were used, while early stopping was applied based on validation loss using a 10% split from the training set. These settings were selected empirically based on preliminary experimentation and are aligned with commonly adopted practices in time series forecasting.

Both deep learning models (LSTM and DNN) were trained using the Adam optimizer with a learning rate of 0.001 and mean squared error (MSE) as the loss function. A batch size of 64 and a maximum of 50 training epochs were used. To prevent overfitting and reduce unnecessary training time, early stopping was applied with a patience of two epochs, monitoring the training loss. A validation split of 10% from the training data was used for this purpose. All models were trained using normalized data sequences, reshaped into appropriate formats: 3D tensors for LSTM and flattened 2D arrays for DNN.

V. Experimental Results

This section presents a comprehensive experimental evaluation of the forecasting framework across multiple datasets, targets, and operational scenarios. The analysis is organized in four stages. First, short-term load forecasting results for the Samos island are reported, including both overall one-step-ahead accuracy on the full test set and a rolling 7-day stress-test evaluation on a highly volatile segment. Next, the performance of the models on the Samos wind generation dataset is examined, highlighting both standard test-set accuracy and the behavior under rolling and recursive multi-step deployment. The third part focuses on load forecasting for the Ikaria island,

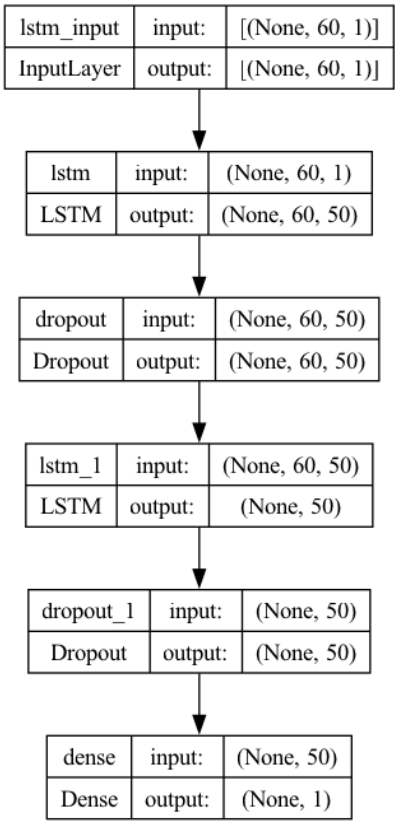


Fig. 1: Compact view of the LSTM forecasting model architecture.

assessing both long-horizon stability and short-term predictive accuracy. Finally, wind power forecasting results for Ikaria are presented under both grid-connected and hybrid operational scenarios.

Across all experiments, we compare classical statistical models (SARIMAX, ARIMA, and AutoRegressive) against deep learning approaches (LSTM and DNN) using RMSE, MAE, and sMAPE metrics. In addition to standard test-set evaluation, special emphasis is placed on realistic operational stress tests, including rolling 1-step-ahead forecasting and recursive multi-step deployment, in order to assess model robustness, error accumulation, and long-horizon stability under real-world conditions.

A. Samos Island — Load Forecasting

This subsection presents the experimental results for short-term load forecasting on the Samos island dataset. The evaluation includes the same methods as the Ikaria's case. Two complementary evaluation protocols are considered: (i) standard one-step-ahead prediction on the full test set, and (ii) a rolling 1-step-ahead forecasting experiment over a challenging 168-step (7-day) horizon selected from a highly volatile segment of the time series.

1) Overall One-Step-Ahead Test Performance: Table I reports the forecasting accuracy of all models on the full test set using RMSE, MAE, and sMAPE metrics.

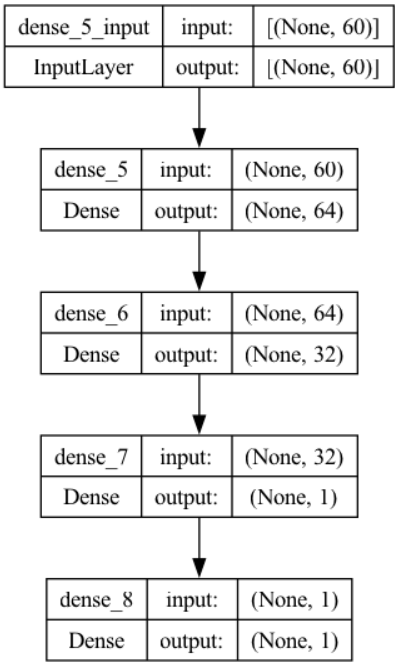


Fig. 2: Compact view of the DNN forecasting model architecture.

TABLE I: Samos Load — Overall One-Step-Ahead Forecasting Performance (Normalized Units)

Model	RMSE	MAE	sMAPE (%)
SARIMAX	0.63697	0.60986	196.92
ARIMA	1.30099	1.15886	179.36
AutoReg	0.20500	0.15428	25.01
LSTM	0.02373	0.01862	2.84
DNN	**0.01597**	**0.01106**	**1.72**

The results clearly demonstrate the superiority of deep learning models over classical statistical approaches. Both LSTM and DNN achieve more than an order-of-magnitude improvement in error compared to ARIMA, SARIMAX, and AutoReg. Among them, the DNN model slightly outperforms LSTM across all metrics, achieving the lowest RMSE, MAE, and sMAPE. This confirms that nonlinear deep models are significantly more capable of capturing the complex temporal dynamics and seasonality patterns present in the Samos load demand time series.

2) Rolling 7-Day Forecast on a Volatile Segment (Operational Stress Test): To evaluate the models under realistic operational conditions, we performed a rolling 1-step-ahead forecast over a 168-step (7-day) horizon, starting from an automatically selected highly volatile segment of the test set (standard deviation ≈ 0.153). Figure 3 illustrates the ground truth and the rolling predictions of both LSTM and DNN in real units.

The quantitative performance over this difficult forecasting interval is:

- **LSTM:** RMSE = 1.3352, MAE = 1.1001, sMAPE =

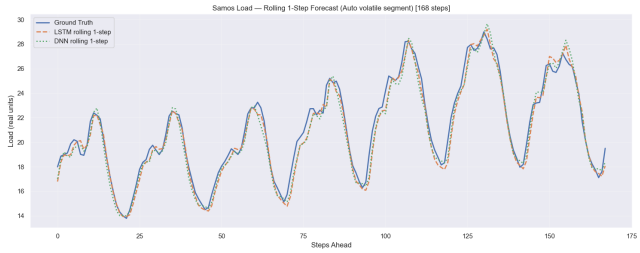


Fig. 3: Samos Load — Rolling 1-step Forecast over 168 Steps on a Volatile Segment

5.29%

- **DNN:** RMSE = 0.6474, MAE = 0.5063, sMAPE = 2.39%

These results confirm that although both models perform well in standard one-step-ahead evaluation, the DNN exhibits substantially better stability and robustness under long-horizon rolling forecasting. The LSTM accumulates noticeably more error over time, whereas the DNN maintains closer tracking of the true load trajectory. This experiment highlights the importance of stress-testing forecasting models under realistic multi-day deployment scenarios, where error accumulation effects become critical.

B. Samos Wind Forecasting Results

Table II reports the quantitative performance of all examined models on the Samos wind dataset. Classical statistical approaches (ARIMA and SARIMAX) exhibit very poor performance, with extremely high sMAPE values close to 200%, indicating their inability to capture the nonlinear and highly volatile dynamics of wind generation. The AutoReg model improves performance but remains significantly inferior to deep learning methods.

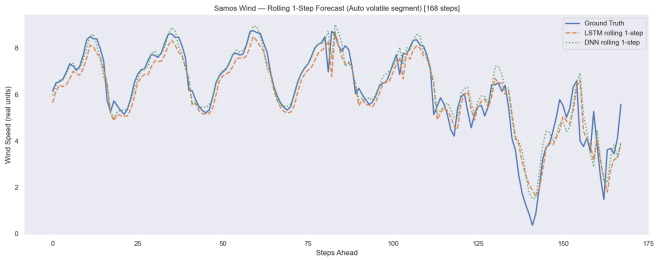
Both neural models achieve a dramatic error reduction. The DNN slightly outperforms the LSTM, achieving the lowest RMSE (0.0592), MAE (0.0374), and sMAPE (9.74%), confirming its superior approximation capability on this dataset.

TABLE II: Performance comparison on the Samos wind dataset (test set).

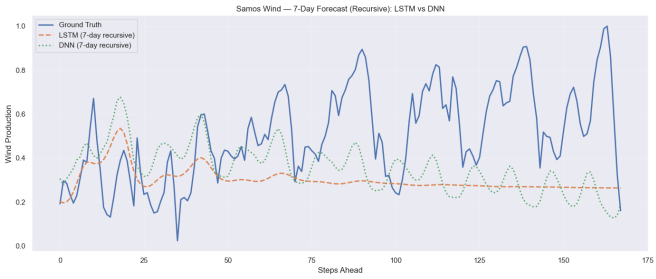
Model	RMSE	MAE	sMAPE (%)
SARIMAX	0.5527	0.4972	199.96
ARIMA	2.6244	2.3771	199.98
AutoReg	0.2537	0.2064	46.23
LSTM	0.0642	0.0465	10.97
DNN	**0.0592**	**0.0374**	**9.74**

Figure 4b presents the rolling 1-step-ahead forecasts over an automatically selected high-volatility segment. Both LSTM and DNN closely track the ground truth, including during abrupt changes, with the DNN exhibiting slightly sharper peak and trough alignment.

Figure 4a shows the 7-day recursive multi-step forecasts. As expected, recursive forecasting leads to error accumulation, particularly for the LSTM, which gradually drifts toward a biased mean trajectory. The DNN demonstrates greater stability and retains partial oscillatory structure, although both models



(a) Rolling 1-step-ahead forecast on an automatically selected volatile segment (168 steps).



(b) 7-day recursive multi-step forecast using LSTM and DNN.

Fig. 4: Samos wind forecasting performance: (a) rolling 1-step-ahead prediction on a highly volatile segment, and (b) 7-day recursive multi-step forecasting.

show degradation at longer horizons. This behavior highlights the intrinsic difficulty of long-horizon wind forecasting under recursive deployment and motivates the use of short-horizon rolling strategies in operational settings.

C. Ikaria Load Forecasting Results

Figure 6 presents the rolling 1-step-ahead forecasting performance on an automatically selected volatile segment of the Ikaria load dataset. Both deep learning models successfully capture the seasonal structure and short-term dynamics of the signal, with the DNN providing slightly tighter tracking of peaks and troughs compared to the LSTM.

TABLE III: Forecasting performance comparison for Ikaria load dataset.

Model	RMSE	MAE	sMAPE (%)
SARIMAX	0.4762	0.4085	198.21
ARIMA	0.2574	0.2189	50.27
AutoReg	0.2438	0.1806	42.85
LSTM	0.0645	0.0405	8.06
DNN	**0.0565**	**0.0370**	**7.82**

Figure 5 illustrates the overall predictive behavior on the full test set, demonstrating stable long-term tracking without significant drift or amplitude collapse.

Quantitative results are summarized in Table III. Classical statistical models (SARIMAX, ARIMA, AutoReg) exhibit substantially higher errors, confirming their limited ability to model the nonlinear and non-stationary behavior of the load signal. In contrast, both neural models achieve an order-of-magnitude lower error.

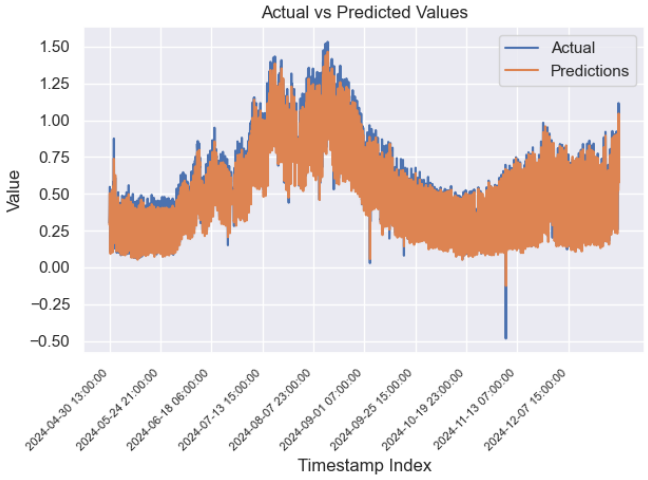


Fig. 5: Ikaria load — full test-set actual vs predicted values (DNN model).

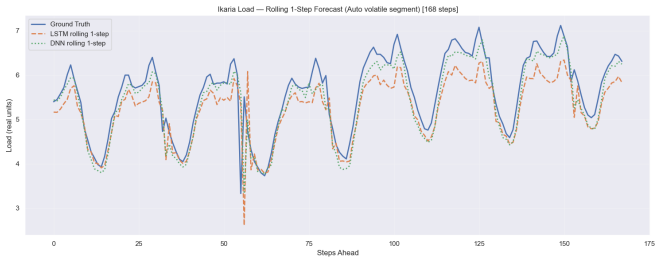


Fig. 6: Ikaria load — rolling 1-step-ahead forecast on an automatically selected volatile segment (168 steps).

The DNN achieves the best overall performance with RMSE = 0.0565, MAE = 0.0370, and sMAPE = 7.82%, slightly outperforming the LSTM. These results confirm the strong suitability of deep learning models for short-term load forecasting in islanded energy systems under highly dynamic conditions.

D. Ikaria Wind Forecasting

Tables IV and V report the forecasting performance for the Ikaria wind power time series under the grid and hybrid potential scenarios, respectively. As in previous cases, all deep learning models significantly outperform the classical statistical approaches.

TABLE IV: Performance comparison for Ikaria wind forecasting (Grid potential).

Model	RMSE	MAE	sMAPE (%)
SARIMAX	0.5101	0.4112	199.50
ARIMA	1.0146	0.8698	196.16
AutoReg	0.3330	0.2491	66.67
LSTM	0.1349	0.0874	34.13
DNN	0.1334	0.0884	36.62

For the grid potential case, both neural models achieve an order-of-magnitude improvement over SARIMAX, ARIMA, and AutoReg. The DNN attains the lowest RMSE (0.1334), while the LSTM follows very closely (0.1349), indicating comparable accuracy. In contrast, ARIMA and SARIMAX exhibit extremely high sMAPE values (approximately 200%), confirming their inability to track the highly non-linear and intermittent dynamics of wind generation.

A similar behavior is observed in the hybrid potential scenario. The LSTM achieves the best overall performance (RMSE = 0.1332, MAE = 0.0860), while the DNN shows slightly degraded accuracy (RMSE = 0.1434). Nevertheless, both neural models remain dramatically superior to all classical baselines, which again display very large relative errors.

TABLE V: Performance comparison for Ikaria wind forecasting (Hybrid potential).

Model	RMSE	MAE	sMAPE (%)
SARIMAX	0.5100	0.4112	199.50
ARIMA	1.0143	0.8695	196.16
AutoReg	0.3330	0.2492	66.68
LSTM	0.1332	0.0860	36.68
DNN	0.1434	0.1035	39.99

Overall, the results demonstrate that deep learning models generalize robustly across both grid and hybrid wind production regimes, while statistical models fail to capture the strong non-stationarity and volatility inherent in wind power time series.

VI. Discussion

The experimental evaluation across the Samos and Ikaria datasets confirms the clear superiority of deep learning models over classical statistical approaches for both load and wind forecasting. In all scenarios, SARIMAX, ARIMA, and AutoReg exhibit significantly higher errors, particularly in wind forecasting, where their sMAPE values approach 200%, indicating their inability to model the strong non-linearity and intermittency of renewable generation.

Both LSTM and DNN achieve consistently low errors across all tasks, with the DNN showing slightly better robustness in most cases, especially under rolling and long-horizon recursive forecasting. While the LSTM performs competitively in one-step-ahead prediction, it exhibits stronger error accumulation in recursive multi-step deployment, whereas the DNN maintains more stable behavior. A clear distinction is observed between load and wind forecasting. Load time series are more regular and predictable, leading to very low errors for both neural models. In contrast, wind power remains intrinsically harder to predict, and all models degrade under long-horizon recursive forecasting, highlighting the need for short-horizon rolling strategies in operational settings. This observation is consistent with recent reviews of deep learning-based wind power forecasting, which emphasize the remaining challenges and future research needs in this area [13].

Finally, the consistent behavior across both islands and across grid and hybrid wind regimes demonstrates that the proposed deep learning pipeline generalizes well and is suitable for deployment in real-world island energy management systems. Table VI summarizes the relative performance of all

TABLE VI: RMSE comparison across all datasets and tasks.

Model	Samos Load	Samos Wind	Ikaria Load	Ikaria Wind (Grid)	Ikaria Wind (Hybrid)
SARIMAX	0.637	0.553	0.476	0.510	0.510
ARIMA	1.301	2.624	0.257	1.015	1.014
AutoReg	0.205	0.254	0.244	0.333	0.333
LSTM	0.024	0.064	0.065	0.135	**0.133**
DNN	**0.016**	**0.059**	**0.056**	**0.133**	0.143

models across all forecasting tasks. The results clearly show that classical statistical models are consistently one to two orders of magnitude worse than deep learning approaches. The DNN achieves the best or near-best performance in all scenarios and exhibits the best average normalized score, confirming its superior robustness and generalization ability.

Despite the strong empirical performance, several limitations should be acknowledged. The proposed models are purely data-driven and do not incorporate exogenous variables such as meteorological forecasts or calendar information, which could further improve predictive performance, particularly for wind power. In addition, long-horizon recursive forecasting remains inherently unstable for highly stochastic processes, and even deep models exhibit error accumulation over time. Finally, this study focuses on point forecasts; extending the framework toward probabilistic forecasting and uncertainty quantification constitutes an important direction for future work.

VII. Conclusion

This paper presented a comprehensive evaluation of short-term load and wind power forecasting in island energy systems using both classical statistical methods and deep learning models. Experiments were conducted on high-resolution datasets from the Samos and Ikaria islands, covering multiple forecasting tasks, including load demand and wind generation under different production regimes. The results consistently demonstrated that deep learning approaches, namely LSTM and DNN, significantly outperform traditional statistical models such as SARIMAX, ARIMA, and AutoReg across all scenarios. In particular, the DNN model exhibited the most robust and stable behavior in the majority of cases, especially under rolling and long-horizon forecasting experiments, while the LSTM showed competitive accuracy but greater sensitivity to error accumulation under recursive deployment.

The experimental analysis also highlighted fundamental differences between load and wind forecasting. While load time series proved highly predictable with very low error levels, wind power forecasting remained intrinsically more challenging due to strong intermittency and non-stationarity, making short-horizon rolling forecasting strategies more suitable for operational deployment. Overall, the proposed deep learning-based forecasting framework provides a reliable and scalable solution for short-term prediction in islanded energy systems and decision-support platforms. Future work will focus on incorporating exogenous information, such as meteorological forecasts, and extending the framework toward probabilistic forecasting and uncertainty-aware decision making.

References

[1] Y. Eren and İ. Küçükdemiral, "A comprehensive review on deep learning approaches for short-term load forecasting," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 114031, 2024.

[2] E. Iturbe, E. Rios, C. Dalamagkas, P. Radoglou-Grammatikis, and N. Toledo, "A pattern-aware lstm-based approach for apt detection leveraging a realistic dataset for critical infrastructure security," *Future Generation Computer Systems*, p. 108308, 2025.

[3] K. Ullah, M. Ahsan, S. M. Hasanat, M. Haris, H. Yousaf, S. F. Raza, R. Tandon, S. Abid, and Z. Ullah, "Short-term load forecasting: A comprehensive review and simulation study with cnn-lstm hybrids approach," *IEEE Access*, 2024.

[4] A. Kumar, A. Singh, and S. Kumar, "Forecasting wind speed over multiple horizons: Superiority of deep learning and decomposition-driven hybrid models," *Unconventional Resources*, p. 100296, 2025.

[5] H. He, Y. Huang, and T. Senjyu, "Microgrid load forecasting and optimization for future remote island electric vehicle mass penetration: The case of okinawa island," *Energy Reports*, vol. 11, pp. 5532–5541, 2024.

[6] C. Dalamagkas, P. Radoglou-Grammatikis, P. Bouzinis, I. Papadopoulos, T. Lagkas, V. Argyriou, S. Goudos, D. Margounakis, E. Fountoukidis, and P. Sarigiannidis, "Federated detection of open charge point protocol 1.6 cyberattacks," *arXiv preprint arXiv:2502.01569*, 2025.

[7] N. Andriopoulos, A. Magklaras, A. Birbas, A. Papalexopoulos, C. Valouxis, S. Daskalaki, M. Birbas, E. Housos, and G. P. Papaioannou, "Short term electric load forecasting based on data transformation and statistical machine learning," *Applied Sciences*, vol. 11, no. 1, p. 158, 2020.

[8] Q. Shen, L. Mo, G. Liu, J. Zhou, Y. Zhang, and P. Ren, "Short-term load forecasting based on multi-scale ensemble deep learning neural network," *IEEE Access*, vol. 11, pp. 111 963–111 975, 2023.

[9] M. López, S. Valero, C. Senabre, and A. Gabaldón, "Analysis of the influence of meteorological variables on real-time short-term load forecasting in balearic islands," in *2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2017, pp. 10–15.

[10] E. Caro and J. Juan, "Short-term load forecasting for spanish insular electric systems," *Energies*, vol. 13, no. 14, p. 3645, 2020.

[11] I. Makris, N. Moschos, P. Radoglou-Grammatikis, N. Andriopoulos, N. Tzanis, K.-N. Malamaki, M. Fotopoulou, K. Kaousias, G. L. Tompros, and P. Sarigiannidis, "Exploring load forecasting: Bridging statistical methods to deep learning techniques in real-world environment," in *2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, 2024, pp. 491–496.

[12] J. Keisler, S. Claudel, G. Cabriel, and M. Brégère, "Automated deep learning for load forecasting," in *Proceedings of the Third International Conference on Automated Machine Learning*, ser. Proceedings of Machine Learning Research, K. Eggensperger, R. Garnett, J. Vanschoren, M. Lindauer, and J. R. Gardner, Eds., vol. 256. PMLR, 09–12 Sep 2024, pp. 16/1–28. [Online]. Available: https://proceedings.mlr.press/v256/keisler24a.html

[13] A. Choudhary, P. Jain, and A. Prajesh, "Wind power forecasting using deep learning method: A review," in *2023 1st International Conference on Intelligent Computing and Research Trends (ICRT)*, 2023, pp. 1–6.