





## Article

# Hybrid AI-Based Framework for Renewable Energy Forecasting: One-Stage Decomposition and Sample Entropy Reconstruction with Least-Squares Regression

Nahed Zemouri <sup>1</sup>, Hatem Mezaache <sup>1</sup>, Zakaria Zemali <sup>2,3</sup> , Fabio La Foresta <sup>3</sup> , Mario Versaci <sup>3</sup>   
and Giovanni Angiulli <sup>4,\*</sup> 

<sup>1</sup> Electrical Engineering Laboratory (LGE), Department of Electronics, Faculty of Technology, University Mohamed Boudiaf M'Sila, M'Sila 28000, Algeria; nehed.zemouri@univ-msila.dz (N.Z.); hatem.mezaache@univ-msila.dz (H.M.)

<sup>2</sup> Laboratory of Applied Automation and Industrial Diagnostics (LAADI), Faculty of Science and Technology, Ziane Achour University, Djelfa 17000, Algeria; z.zemali@univ-djelfa.dz

<sup>3</sup> Department of Civil, Energetic, Environmental and Material Engineering, Mediterranea University, I-89124 Reggio Calabria, Italy; fabio.laforesta@unirc.it (F.L.F.); mario.versaci@unirc.it (M.V.)

<sup>4</sup> Department of Information Engineering, Infrastructures and Sustainable Energy, Mediterranea University, I-89124 Reggio Calabria, Italy

\* Correspondence: giovanni.angiulli@unirc.it

**Abstract:** Accurate renewable energy forecasting is crucial for grid stability and efficient energy management. This study introduces a hybrid model that combines signal decomposition and artificial intelligence to enhance the prediction of solar radiation and wind speed. The framework uses a one-stage decomposition strategy, applying variational mode decomposition and an improved empirical mode decomposition method with adaptive noise. This process effectively extracts meaningful components while reducing background noise, improving data quality, and minimizing uncertainty. The complexity of these components is assessed using entropy-based selection to retain only the most relevant features. The refined data are then fed into advanced predictive models, including a bidirectional neural network for capturing long-term dependencies, an extreme learning machine, and a support vector regression model. These models address nonlinear patterns in the historical data. To optimize forecasting accuracy, outputs from all models are combined using a least-squares regression technique that assigns optimal weights to each prediction. The hybrid model was tested on datasets from three geographically diverse locations, encompassing varying weather conditions. Results show a notable improvement in accuracy, achieving a root mean square error as low as 2.18 and a coefficient of determination near 0.999. Compared to traditional methods, forecasting errors were reduced by up to 30%, demonstrating the model's effectiveness in supporting sustainable and reliable energy systems.

**Keywords:** solar radiation; wind speed; hybrid deep learning forecasting; least-squares regression hybrid model



check for updates

Academic Editor: John Boland

Received: 1 May 2025

Revised: 26 May 2025

Accepted: 27 May 2025

Published: 3 June 2025

**Citation:** Zemouri, N.; Mezaache, H.; Zemali, Z.; La Foresta, F.; Versaci, M.; Angiulli, G. Hybrid AI-Based Framework for Renewable Energy Forecasting: One-Stage Decomposition and Sample Entropy Reconstruction with Least-Squares Regression. *Energies* **2025**, *18*, 2942. <https://doi.org/10.3390/en18112942>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent decades, the transition towards a sustainable energy system has become a global priority, driven by the need to reduce greenhouse gas emissions and limit dependence on fossil fuels [1,2]. In this context, renewable energy sources—particularly solar [3] and wind [4]—play a key role in the decarbonization process [5,6]. However, due to their intermittent nature and strong dependence on meteorological variables, energy

production becomes a complex challenge [7]. The accurate forecasting of future generation is essential to ensure grid stability, optimize demand-side management, and reduce operational costs [8].

The scientific literature has recently proposed numerous models for renewable energy forecasting [9], ranging from traditional statistical techniques to artificial intelligence-based methods [10,11]. Physical models rely on meteorological and topographical data to estimate energy output, but their applicability is limited by the complexity of atmospheric phenomena and the requirement for high-resolution data [12,13]. On the other hand, statistical and autoregressive models effectively capture historical patterns but prove inadequate when faced with nonlinear dynamics and sudden meteorological changes [14,15]. The introduction of machine learning techniques and neural networks has revolutionized the field, enabling the modeling of complex relationships between weather variables and energy production [16]. However, their effectiveness is often hindered by noise in the data and the inability to distinguish informative components from irrelevant ones accurately [17,18].

In this context, the present study introduces an innovative framework that overcomes the limitations of existing models through an advanced integration of signal decomposition techniques, entropy-based feature selection, and predictive models based on deep learning and least-squares regression. The main innovation here compared to the recent literature lies in the combined use of variational mode decomposition (VMD) and improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) for data pre-processing—two of the most advanced techniques for analyzing non-stationary time series. These methods enable the decomposition of historical solar radiation and wind speed data into more interpretable modal components, thereby enhancing input data quality and reducing error propagation in predictive models.

Unlike conventional models that employ multi-stage decomposition strategies with high computational complexity [19–21], the proposed framework adopts a single-stage decomposition approach, ensuring improved computational efficiency without compromising forecast accuracy. An additional innovative feature is the use of sample entropy [22,23] for automatically selecting the most relevant components, allowing for the isolation of key information and the more effective filtering of residual noise compared to traditional methods. This phase enhances the model's ability to detect significant variations in the data and adapt more effectively to dynamic weather conditions.

The proposed framework thus integrates advanced predictive models, including bidirectional recurrent neural networks [24,25], extreme learning machines, and support vector regression, all optimized to model nonlinear dynamics and capture complex relationships within energy data [26,27]. The final forecast is obtained through a least-squares regression-based hybrid strategy, which optimally combines the contributions of the different models, reducing prediction error and improving system stability. Compared to traditional approaches, this methodology enables more accurate and robust forecasts, as demonstrated by experimental results on real-world data from three distinct locations characterized by heterogeneous climatic conditions.

A hybrid deep learning model combining CNN and LSTM was used to predict solar radiation, where the results showed an improvement in predicting radiation over various time intervals [28]. In this study, a new LSTM model with adaptive wind speed calibration (C-LSTM) was presented, which significantly improved the accuracy of predictions. This was applied to 25 wind turbines by dynamically adjusting the expected wind speed during all phases of training and inference throughout the study [29]. To improve the performance of the LSTM training process, the AROA algorithm was used for this purpose, where the results showed a significant improvement in the accuracy of wind speed predictions, and the study was conducted using real wind speed data [30]. In this study, a hybrid model for

deep learning was used employing the LSTM algorithm, which was optimized through hyperparameter tuning. The results showed a significant improvement in short-term solar radiation prediction [30].

The comparative analysis conducted in this study highlights a significant improvement over existing models, with reduced forecasting errors and enhanced generalization capability. In particular, the developed hybrid model achieves superior performance in terms of prediction accuracy, with error values being reduced by up to 30% compared to conventional techniques and a coefficient of determination approaching 0.999, demonstrating the validity of the proposed approach.

The structure of the paper is organized as follows: Section 2 provides an overview of the recent scientific literature on renewable energy forecasting techniques and signal decomposition methodologies. Section 3 describes the proposed methodology, focusing on pre-processing techniques, machine learning models, and the forecast hybrid strategy. Section 4 offers a detailed description of the proposed approach, while Section 5 presents the experimental setup and the datasets used for model validation. Subsequently, Section 6 provides insights into the computational complexity and the proposed model's efficiency evaluation. The experimental results are thoroughly discussed in Section 7, with Section 8 being devoted to concluding remarks and potential future developments of the ongoing research.

## 2. Related Works

Physical models use historical meteorological and topographical data [31–36], while statistical models forecast wind speed and solar irradiance [37] by modeling time series using autoregressive approaches (AR [38], MA [39], and ARMA [40]). In [41], an adaptive model for Numerical Weather Prediction (NSD) demonstrated that error correction significantly improves the accuracy of weather forecasts.

Statistical models predict wind speed and solar irradiance using autoregressive approaches (AR, MA, ARMA, ARIMA, and SARIMA) [42], tuning parameters and comparing predicted values with actual measurements. ARMA models are prevalent, as they effectively represent wind speed characteristics while minimizing error [43,44].

Another class of forecasting methods for wind speed and solar irradiance includes machine learning (ML) techniques [45], which have enabled the development of various accurate forecasting models, including artificial neural networks (ANNs) [46], Adaptive Neuro-Fuzzy Inference Systems (ANFISs) [47–51], complex neuro-fuzzy networks, and innovative approaches based on Support Vector Machines (SVMs) [52]. Specifically, an ANN model was developed in [53] to predict wind speed, leveraging local meteorological data to learn nonlinear features and improve forecast accuracy. In [43], a model based on Back Propagation Neural Networks (BPNN) and time series was used for short-term solar irradiance forecasting, optimizing the network through cross-validation and parameter tuning to prevent overfitting and enhance predictive precision. In [44], two models for wind speed forecasting were proposed: one based on ANNs and the other on ANFISs, using different network configurations and meteorological measurements. The results showed superior accuracy for the ANFIS model. In [54], an advanced model combining SVMs and k-means clustering was proposed to forecast solar irradiance one hour ahead, improving forecasting accuracy and optimizing the performance of Energy Storage Systems (ESSs).

Hybrid approaches combine linear and nonlinear models to improve short- and long-term forecasts. In [53], a multi-step architecture for wind speed forecasting was developed, integrating decomposition algorithms and modified ANNs. Among the four tested hybrid models, the one based on Singular Spectrum Analysis (SSA) and a General Regression Neural Network (GRNN) yielded the most accurate forecasts. The evolution of artifi-

cial intelligence (AI), machine learning, and deep learning has made these techniques essential for accurately forecasting nonlinear phenomena such as wind speed and solar irradiance. In [55], a hybrid approach was proposed based on empirical mode decomposition (EMD) and its advanced versions—ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)—combined with Singular Spectrum Analysis (SSA) for processing high-frequency components and the Elman Neural Network (ENN). It is worth noting that in [56], a hybrid model for short-term wind speed forecasting was proposed, combining CEEMDAN and variational mode decomposition (VMD) with an advanced AdaBoost algorithm and an extreme learning machine (ELM). This model addresses the nonlinearity of wind speed time series and improves forecast accuracy. The approach proposed in [57] employs the Seasonal Autoregressive Integrated Moving Average (SARIMA) model—derived from the ARIMA model—to forecast wind speed and solar irradiance [37,58]. SARIMA, designed to handle seasonality in time series, follows a four-step process (identification, estimation, diagnostic checking, and forecasting) to optimize the design and sizing of hybrid renewable energy systems, combining wind, solar, and storage technologies to ensure reliable energy supply. A summary of the main forecasting approaches discussed in the literature, along with their respective strengths and limitations, is reported in Table 1, which also highlights the motivation for developing the proposed hybrid method.

**Table 1.** Summary of selected representative studies on renewable energy forecasting, including the forecasting methods used, main advantages and limitations, and the research gaps identified. This comparison provides context for the development of the proposed hybrid framework.

Model/Study	Technique	Advantages	Limitations
ARIMA/SARIMA-based models [43,44]	Statistical time series modeling	Interpretable, efficient for linear patterns	Poor performance on nonlinear and abrupt changes
ANN/BPNN [46,53]	Neural networks	Learns nonlinearities, adaptable	Risk of overfitting, requires parameter tuning
ANFIS [46–51]	Neuro-fuzzy system	Handles uncertainty, interpretable rules	Sensitive to noise, computationally intensive
SVM + k-means [54]	ML with clustering	Enhances solar irradiance forecasting	Requires pre-segmentation, lacks adaptability
SSA + GRNN [53]	Hybrid model with decomposition	Improves forecast accuracy	Increased complexity and cost
CEEMDAN + AdaBoost + ELM [56]	Advanced hybrid decomposition + ensemble learning	Captures nonlinearity effectively	High computational burden, no entropy selection
Proposed model	VMD/ICEEMDAN + SE + BiLSTM/ELM/SVR + LSR	Integrates decomposition, entropy-based filtering, and ensemble learning; high accuracy and efficiency	–

### 3. Overview of the Forecasting Strategy and System Architecture

This study proposes an advanced hybrid framework for the short-term forecasting of two renewable energy sources (RESs): global horizontal irradiance (GHI) and wind speed (WS). The model is based on a multilayer predictive architecture that integrates signal decomposition techniques, entropy measures, and machine learning models to improve forecast accuracy and robustness.

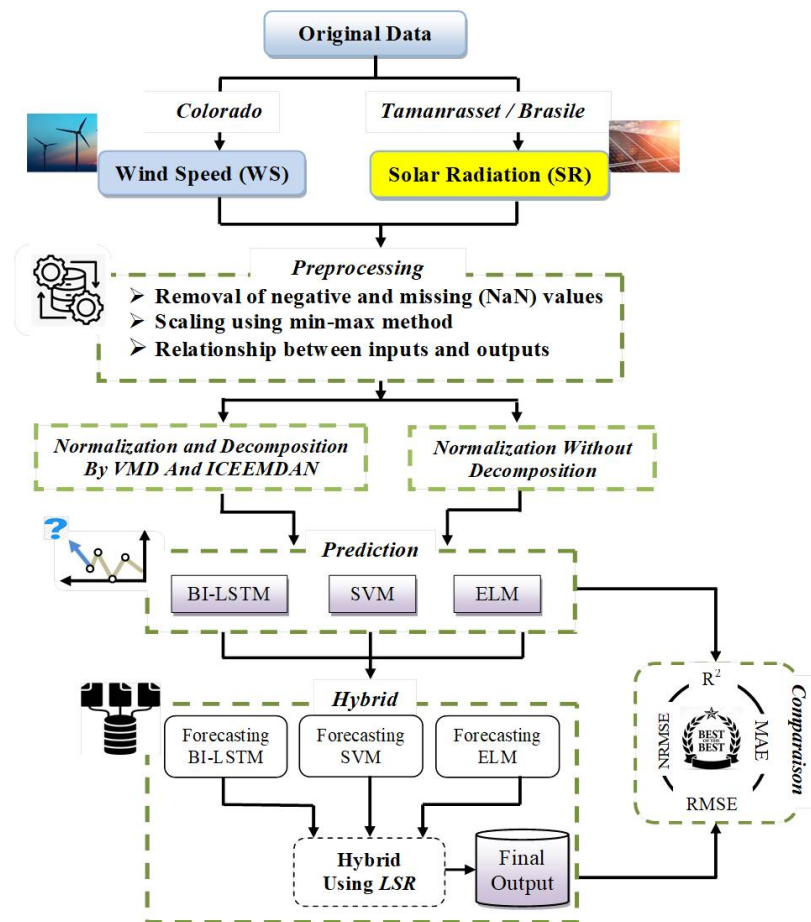
The main innovation lies in the combined use of variational mode decomposition (VMD), improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), and sample entropy (SampEn), applied within a single-stage decomposition strategy. This combination allows for a more effective separation of informative components from noise compared to traditional methods, improving the quality of the input data for the predictive models.

The forecasting process consists of several steps:

1. Signal decomposition: VMD and ICEEMDAN decompose GHI and WS signals into modal subcomponents with distinct frequency bands. VMD is configured with 12 modes and a bandwidth constraint of 2000, while ICEEMDAN uses 100 iterations and a white noise standard deviation of 0.2.
2. Entropy-based selection: sample entropy is used to identify the most informative subcomponent, removing redundant or noisy components. This step reduces computational complexity and enhances the quality of the reconstructed signal.
3. Predictive modeling: Each significant subcomponent is processed by three distinct models:
  - Bidirectional Long Short-Term Memory (Bi-LSTM), effective at capturing long-term dependencies in both temporal directions;
  - Extreme learning machine (ELM), a single-layer feedforward network with randomly assigned weights, offering high training speed;
  - Support vector regression (SVR), which models complex nonlinear relationships with strong generalization capability.
4. Model hybrid: The predictions from the individual models are combined through a least-squares regression (LSR) strategy, which assigns optimal weights to each model based on its performance, enhancing the stability and accuracy of the final output.

The framework was tested on real-world datasets from three locations with heterogeneous climatic conditions: Tamanrasset (Algeria), Brasilia, and Colorado (USA). The results show an error reduction of up to 30% compared to conventional models and a coefficient of determination ( $R^2$ ) close to 0.999, demonstrating the effectiveness and reliability of the proposed method. Furthermore, the framework was evaluated in terms of computational complexity, showing that the VMD-based approach offers an optimal trade-off between accuracy and processing time (18–35 s for 10,000 samples), making it suitable for real-world applications and operational deployment. The flowchart of the forecasting strategy, depicted in Figure 1, provides a visual representation of the developed model and helps to understand the system's overall functioning. Particularly, Figure 1 in the paper illustrates the overall flow of the developed model for forecasting solar radiation and wind speed. It visually represents the system architecture, starting from the transformation of data vectors into matrices and the creation of a test vector, denoted as  $Y_{test}$ , which is later used to generate the forecasts. Figure 1 outlines two scenarios in the preprocessing phase: the first, which is more comprehensive, involves the joint application of the ICEEMDAN and VMD decomposition techniques; the second, simpler scenario uses the  $Y_{test}$  vector directly without any decomposition. The section of Figure 1 depicting the flow with ICEEMDAN and VMD actually corresponds to the implementation of the one-stage decomposition, a distinctive feature of the proposed model compared to many existing multi-stage strategies. In this step, the original signal is divided into modal components through a single joint operation using both techniques, enhancing computational efficiency without sacrificing forecasting accuracy. Following the decomposition, although not explicitly represented in the figure, lies a crucial phase of the process: entropy-based selection. This step, which logically occurs immediately after the decomposition shown in the upper left part of the figure, serves to identify the informative subcomponents of the signal. The selected components are those exhibiting significant complexity according to the sample entropy measure, while redundant or noise-dominated components are discarded. This not only improves the quality of the reconstructed signal but also reduces the computational load, ensuring that only relevant information is passed on to the predictive models. In the diagram, this selection step would ideally be located between the decomposition block and the forecasting modules, but it is not visually depicted. Finally, the signals reconstructed from the

selected informative components are processed by the predictive models shown in the central right section of Figure 1, namely Bi-LSTM, SVM, and ELM. The results produced by these models are ultimately combined through a least-squares regression-based strategy, represented as the final stage in the diagram, which yields the integrated forecast output. In conclusion, although Figure 1 provides a helpful overview of the entire process, it omits key elements such as the entropy-based selection and does not visually clarify the implementation of the one-stage decomposition—both of which are central to fully understanding the methodological innovations of the model.



**Figure 1.** General architecture of the proposed hybrid forecasting framework. The system is applied to solar radiation (SR) prediction for Temenrasset (Algeria) and Brasilia (Brazil), and wind speed (WS) prediction for Colorado (USA). The input data vectors are first transformed into matrix form and split into training and test sets ( $Y_{test}$ ). The preprocessing phase includes a comparison between two configurations: (a) a full pipeline employing ICEEMDAN and VMD decomposition, followed by Variance Mode Filtering (VMF) and sample entropy (SE) for informative subcomponent selection; and (b) a simplified baseline using the original test vector without any decomposition. The forecasting phase integrates three predictive models—Bi-LSTM, ELM, and SVM—whose outputs are combined via least-squares regression (LSR) to produce the final forecast.

**Remark 1.** The proposed forecasting framework follows a sequential and modular architecture designed to enhance both accuracy and interpretability. The process begins with the input of raw time series data, such as global horizontal irradiance (GHI) or wind speed (WS), collected from real-world sensors. These input signals are often non-stationary and contain noise or redundant patterns, which can negatively impact model performance. To address this, a one-stage signal decomposition phase is applied using either variational mode decomposition (VMD) or improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN). This step separates the original

signal into several intrinsic mode functions (IMFs), each capturing distinct frequency components, thereby improving data structure and reducing noise. Following decomposition, a sample entropy (SE)-based selection step is introduced. The SE value of each sub-signal is computed, and only those components with higher entropy—indicating greater informational content and reduced redundancy—are retained. This allows for dimensionality reduction and prevents low-informative or noisy sub-signals from being passed to the prediction models. The selected components are then processed in parallel by three different predictive models: Bidirectional Long Short-Term Memory (Bi-LSTM), extreme learning machine (ELM), and support vector regression (SVR). Each model is tailored to capture specific patterns within the data—Bi-LSTM models temporal dependencies, ELM provides high-speed learning, and SVR captures nonlinear relationships. Finally, the outputs of these models are integrated through a hybrid fusion strategy based on least-squares regression (LSR). The LSR model assigns optimal weights to the individual predictions, producing a single, unified output that leverages the strengths of each model while mitigating their individual weaknesses. This final prediction represents the most accurate and stable estimate of the target variable.

**Remark 2.** The proposed forecasting architecture distinguishes itself from conventional approaches through its integrated, single-stage design that combines signal decomposition, entropy-based component selection, and multi-model fusion. Unlike traditional methods that often rely on multi-step or layered decomposition strategies—each introducing additional complexity and error propagation—the present framework employs a unified decomposition step using VMD or ICEEM-DAN. This choice ensures computational efficiency while preserving signal integrity. Furthermore, by incorporating sample entropy (SE) as a selection criterion, the model is able to systematically identify and retain only the most informative sub-signals, effectively reducing the influence of noise and redundancy without manual tuning. This pre-processing enhancement significantly improves the input quality for machine learning models, which are often sensitive to irrelevant features. In contrast to single-model predictors that may struggle to capture all dynamics of complex environmental data, our hybrid architecture leverages the complementary strengths of Bi-LSTM, ELM, and SVR. These models are combined through a least-squares regression (LSR) strategy, which assigns adaptive weights based on performance, resulting in more accurate and robust predictions. Compared to the existing literature, which typically isolates model training from signal preprocessing or relies on static model selection, the proposed framework introduces a dynamic, data-driven mechanism for both input refinement and model integration. This holistic design leads to a reduction in forecasting error (up to 30%) and high coefficients of determination (up to  $R^2 = 0.999$ ), confirming its superior performance across multiple datasets and climate conditions.

## 4. Methodology

### 4.1. Decomposition Methods

#### 4.1.1. Variational Mode Decomposition (VMD)

VMD is an advanced technique that decomposes a complex signal into subcomponents with specific frequency bands, optimizing them simultaneously to ensure stable and precise separation [59]. This method, particularly effective for nonlinear and non-stationary signals, reduces mode mixing and enhances data quality in forecasting models. It finds application in various domains, such as signal processing and renewable energy forecasting [60].

Mathematically, VMD can be formulated as a variational optimization problem aimed at decomposing a signal,  $f(t)$ , into a set of  $k$  intrinsic mode components,  $u_k(t)$ , each associated with a specific and non-overlapping frequency band. The optimization is based on minimizing the total energy of the demodulated modal components, using the Hilbert transform to obtain an analytic representation. A central frequency  $\omega_k$  is defined for each mode, and frequency translation is applied through exponential modulation. The optimization problem is formalized as follows [59,60]:

$$\min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \left\| \frac{\partial}{\partial t} \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \quad (1)$$

subject to the signal reconstruction constraint,  $\sum_{k=1}^K u_k(t) = f(t)$ . In Equation (1),  $u_k(t)$  represents the intrinsic mode components to be extracted, while  $\omega_k$  denotes the central frequencies associated with each component. Moreover,  $\delta(t) + \frac{j}{\pi t}$ , where  $\delta(t)$  is the Dirac delta function, serves as the kernel of the Hilbert transform used to extract the analytic part of the signal [61], and  $e^{-j\omega_k t}$  is the modulation that shifts the frequency of each mode to its assigned center [59,60].

#### 4.1.2. VMD: Key Details

VMD is an advanced technique to decompose a complex signal into modal components, each corresponding to a specific frequency band [59,60]. This method allows for separating different features present in the signal, improving the quality of analysis and facilitating the identification of underlying structures. The effectiveness of the decomposition depends on two main parameters: the number of modes, which determines how many components are extracted from the original signal, and the quadratic penalty term, which affects the quality of the separation between modes [62]. These parameters are crucial in optimizing the decomposition, as an inappropriate choice can lead to inaccurate segmentation with overlapping or excessively fragmented modes [59,60,62].

To achieve effective decomposition, VMD incorporates a bandwidth constraint that regulates the frequency distribution among the extracted modes [59,60,62]. This parameter prevents undesirable overlaps and ensures the clear separation of the information within the signal. Another key element is the Lagrangian stop multiplier, a parameter that influences the stopping criterion of the iterative process [59,60,62]. This value is determined by a regulation factor governing the decomposition's convergence. In some configurations, VMD can be set up to avoid applying additional penalties on the deviation of modes during iterations, allowing the model to naturally adapt to the signal's structure without artificial convergence constraints [59,60,62].

The method excludes a DC component (i.e., a zero-frequency component), meaning any constant mean value in the signal is removed. This approach allows the focus to remain solely on dynamic variations, ensuring a more effective analysis of temporal changes. In addition, the central frequencies of the modes are initialized uniformly, contributing to a balanced distribution of components across the frequency spectrum. A tolerance level is maintained to control the accuracy of mode separation, ensuring the decomposition is effective without introducing significant numerical errors [59,60,62].

VMD is applied to forecasting global solar radiation and wind speed by splitting the signal into a set of distinct modes, including a final residual. This decomposition process enables the isolation of different frequency components, simplifying the identification of patterns useful for prediction [59,60,62]. Due to its ability to effectively handle the non-stationary characteristics of time series, VMD is particularly well suited for forecasting phenomena that exhibit temporal variability. Its application reduces the influence of noise and chaotic fluctuations, improving the precision and reliability of predictive models. The optimized frequency separation and uniform distribution of modes make VMD a fundamental tool for analyzing and forecasting complex energy-related data, such as those associated with solar and wind energy production [59,60,62].

**Remark 3.** Equation (1) can be solved by applying Lagrange multipliers, resulting in a constrained optimization problem that can be addressed using the Augmented Lagrangian method [63]. This approach introduces a penalty term,  $\alpha$ , which controls the penalty on the reconstruction fidelity, along with a Lagrange multiplier  $\lambda(t)$ , leading to the following cost function:

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \frac{\partial}{\partial t} \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle \end{aligned} \quad (2)$$

The solution is obtained iteratively by updating  $u_k$ ,  $\omega_k$ , and  $\lambda$  until convergence is reached.

This formulation effectively separates modal components with distinct frequencies, avoiding mode mixing issues typical of other techniques, such as empirical mode decomposition (for details, see [64]).

#### 4.1.3. Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN)

ICEEMDAN is an advanced version of CEEMDAN, developed to improve the stability of the decomposition, reduce mode mixing, and ensure a clearer separation of the intrinsic mode functions (IMFs) [65]. Compared to EEMD, this algorithm more effectively addresses the main challenges encountered in empirical decompositions, optimizing the overall quality of the results. The technique allows for the decomposition of a time series signal into a finite number of components, effectively managing adaptive noise and correcting frequency aliasing issues. ICEEMDAN introduces a normalization mechanism and a more efficient distribution of noise energy, enhancing the quality of the IMFs by integrating white noise modes into the original signal and effectively reducing residual noise. This method ensures a more stable and robust decomposition process through improved noise handling across iterations. The lower envelope is assumed to be zero at all points. Once the local maxima  $e_{\text{sup}}(t)$  and minima  $e_{\text{inf}}(t)$  are identified, the upper and lower envelopes are constructed by interpolating the maxima and minima using cubic splines. The average of the envelopes is then computed as  $m_1(t) = 0.5 \cdot (e_{\text{sup}}(t) + e_{\text{inf}}(t))$ , and this is subtracted from the original signal to obtain  $h_1(t) = x(t) - m_1(t)$ . If  $h_1(t)$  satisfies the IMF conditions, it is assigned as  $\text{IMF}_1(t)$ ; otherwise, the process is repeated using  $h_1(t)$  as the new signal until convergence, resulting in  $\text{IMF}_1(t) = h_k(t)$ . This IMF is then subtracted from the signal to obtain the first residual:  $r_1(t) = x(t) - \text{IMF}_1(t)$ . The process is repeated until the final residual  $r_n(t)$  becomes monotonic or shows no significant oscillations. Finally, the signal  $x(t)$  can be expressed as  $x(t) = \sum_{k=1}^K \text{IMF}_k + r_K$ , where  $K$  is the total number of extracted IMFs.

To formally describe the ICEEMDAN method, Gaussian white noise  $w_i(t)$  (with unit variance to enhance decomposition robustness) is added to the original signal before extracting the first IMF (thus avoiding mode mixing), resulting in  $x_i(t) = x(t) + \epsilon_0 w_i(t)$ , for  $i = 1, \dots, N$ , where  $\epsilon_0$  is the noise scaling coefficient. Applying EMD to each perturbed signal yields the first IMF as  $\text{IMF}_1(t) = \frac{1}{N} \sum_{i=1}^N \text{EMD}(x_i(t))$ .

After extracting  $\text{IMF}_1(t)$ , it is subtracted from the original signal to compute the first residual:  $r_1(t) = x(t) - \text{IMF}_1(t)$ . ICEEMDAN introduces a dynamic noise adjustment mechanism to maintain decomposition stability. The amount of noise added at each iteration is adjusted based on the energy of the extracted IMF: if  $\gamma_k$  is a regulation coefficient, then  $\epsilon_k = \gamma_k \frac{\|r_k(t)\|}{\|w_i(t)\|}$ . To extract  $\text{IMF}_2(t)$ , adaptive white noise is added to the residual,  $r_{1i}(t) = r_1(t) + \epsilon_1 w_i(t)$ , for  $i = 1, \dots, N$ , resulting in  $\text{IMF}_2(t) = \frac{1}{N} \sum_{i=1}^N \text{EMD}(r_{1i}(t))$ , and the second residual is computed as  $r_2(t) = r_1(t) - \text{IMF}_2(t)$ . This process con-

tinues until a residual that contains no significant oscillations is obtained. The algorithm stops when  $r_k(t)$  can no longer be decomposed into meaningful IMFs, and the original signal is reconstructed as the sum of all extracted IMFs plus the final residual, i.e.,  $x(t) = \sum_{k=1}^K \text{IMF}_k(t) + r_K$  [65,66].

**Remark 4.** ICEEMDAN adopts an adaptive noise adjustment strategy to improve the separation of IMFs, preventing the propagation of artifacts into subsequent components. This approach reduces mode mixing, ensures more accurate decomposition, and provides stability in frequency representation. Furthermore, the normalization of noise energy balances accuracy with reducing residual noise. The method is applicable in various fields, including weather forecasting, fault diagnosis, biomedical signal processing, and financial analysis.

#### 4.2. Sample Entropy

Sample entropy (SE) is a measure of the complexity of a time series [67,68]. It is used to quantify a signal's repetitiveness and unpredictability level. Unlike Approximate Entropy (ApEn), SE does not consider data autocorrelation and omits self-matches, making it more robust and less dependent on the length of the time series [69].

Formally, let  $X = \{x_1, x_2, \dots, x_N\}$  be a time series of length  $N$ . The method first constructs  $m$  embedding vectors,

$$X_i^m = \{x_i, x_{i+1}, \dots, x_{i+m-1}\}, \quad 1 \leq i \leq N - m + 1, \quad (3)$$

which represent subsequences of the signal of length  $m$ . Typically,  $m$  is set to small values, such as 2 or 3, since higher dimensions may lead to unstable estimates due to the scarcity of similar data.

Once the maximum distance between vectors  $X_i^m$  and  $X_j^m$  is computed as

$$d(X_i^m, X_j^m) = \max_{k=1, \dots, m} |x_{i+k-1} - x_{j+k-1}|, \quad (4)$$

the number of similar pairs is counted, which represents the largest difference between the corresponding components of the two vectors. In other words, by defining a similarity threshold  $r$ , usually set as a fraction of the standard deviation of the time series,

$$r = \alpha \cdot \text{std}(X), \quad 0.1 \leq \alpha \leq 0.2, \quad (5)$$

the number of pairs  $(i, j)$ , with  $i \neq j$ , satisfying  $d(X_i^m, X_j^m) \leq r$ , is estimated. The correlation function for dimension  $m$  is then defined as follows:

$$B^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \frac{\text{number of } X_j^m \text{ similar to } X_i^m}{N - m}. \quad (6)$$

The same procedure is then repeated for  $m + 1$ , yielding

$$B^{m+1}(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \frac{\text{number of } X_j^{m+1} \text{ similar to } X_i^{m+1}}{N - m - 1}, \quad (7)$$

from which

$$\text{SE}(m, r, N) = -\ln\left(\frac{B^{m+1}(r)}{B^m(r)}\right). \quad (8)$$

If the value of  $B^{m+1}(r)$  relative to  $B^m(r)$  is low, the time series is highly predictable and regular, resulting in a low SampEn value. Conversely, a high SampEn indicates a more chaotic or random system [23].

**Remark 5.** It is worth noting that  $r$  is a crucial parameter, as it defines the pairs of similar vectors, while a time series that is too short may compromise the reliability of the  $SE(m, r, N)$  estimate, leading to distorted or unrepresentative results of the system's complexity.

**Remark 6.** In the proposed framework,  $SE$  is employed solely during the pre-processing phase as a criterion for selecting the most informative sub-signals obtained from the VMD and ICEEMDAN decomposition processes. Specifically, only the components with higher  $SE$  values—indicating greater complexity and lower regularity—are retained for forecasting. This selection aims to eliminate noise-dominated or redundant components that may degrade model performance. It is important to note that  $SE$  is not used as a metric to evaluate the quality of the predictions or the forecasting models themselves. Therefore, no correlation analysis between entropy values and prediction errors per mode has been carried out. Such analysis, although beyond the scope of this study, represents an interesting research direction for future investigations aimed at better understanding the relationship between input complexity and forecast accuracy.

### 4.3. Forecasting Models

In this section, we present the different forecasting models, Bi-LSTM, ELM, and SVM, used in this study to estimate GHI and WS.

#### 4.3.1. Bidirectional Long Short-Term Memory (Bi-LSTM)

It is an advanced extension of LSTM (Long Short-Term Memory), a recurrent neural network designed to capture long-term dependencies [70–72]. Bi-LSTM enhances the standard LSTM by processing information in both directions, providing a more comprehensive understanding of the context. It consists of two independent LSTM layers: the first is a forward LSTM  $\vec{h}_t$  that processes the sequence in chronological order, from  $t = 1$  to  $t = T$ ; the second is a backward LSTM  $\overleftarrow{h}_t$ , which processes the sequence in reverse order, from  $t = T$  to  $t = 1$  [70–72]. The output of the Bi-LSTM at a given time  $t$  results from the concatenation of the forward and backward hidden states, forming a vector twice the size of the originals, i.e.,

$$h_t = \text{concat}(\vec{h}_t, \overleftarrow{h}_t) = \begin{bmatrix} \vec{h}_t \\ \overleftarrow{h}_t \end{bmatrix}, \quad (9)$$

meaning that if  $\vec{h}_t$  and  $\overleftarrow{h}_t$  are vectors of dimension  $d$ , the new concatenated hidden state  $h_t$  will have dimension  $2d$ , enabling the model to retain information from both directions of the sequence [70–72].

To understand how Bi-LSTM works, we first define the dynamics of a unidirectional LSTM. At each time step  $t$ , the model first computes the input gate as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \quad (10)$$

which regulates the amount of new information to be added to memory, where  $i_t$  is the input gate value at time  $t$ ,  $\sigma(\cdot)$  is the sigmoid activation function,  $W_i$  is the weight matrix for the input vector  $x_t$ ,  $U_i$  is the weight matrix for the previous hidden state  $h_{t-1}$ , and  $b_i$  is the bias term for the input gate [70–72].

Next, the forget gate is computed as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \quad (11)$$

which decides which information from the previous memory should be retained or discarded. Here,  $f_t$  is the forget gate value at time  $t$ , and  $W_f$ ,  $U_f$ , and  $b_f$  are the corresponding weight matrices and bias term [70–72].

The output gate is then calculated as follows:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (12)$$

which determines what part of the memory state contributes to the network's output. In this equation,  $o_t$  is the output gate value at time  $t$ , and  $W_o$ ,  $U_o$ , and  $b_o$  are the associated weights and bias.

These three gates allow the LSTM to manage information over time, optimizing the capture of long-term dependencies in sequences. The candidate memory state is then computed as follows:

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c), \quad (13)$$

which represents a proposed update to the cell state at time  $t$ , calculated using the tanh activation function, compressing values into the interval  $(-1, 1)$ . Here,  $\tilde{C}_t$  is the candidate memory state at time  $t$ , and  $W_c$ ,  $U_c$ , and  $b_c$  are the weight matrices and bias [70–72].

This candidate is combined with  $i_t$  and  $f_t$  to update the cell memory state:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (14)$$

where  $\odot$  denotes the Hadamard (element-wise) product. The final visible output from the cell at time  $t$  is then

$$h_t = o_t \odot \tanh(C_t). \quad (15)$$

The Bi-LSTM uses two independent LSTM networks. The first,

$$\vec{h}_t = \text{LSTM-forward}(x_t, \vec{h}_{t-1}, \vec{C}_{t-1}), \quad (16)$$

processes the data forward, while the second,

$$\overleftarrow{h}_t = \text{LSTM-backward}(x_t, \overleftarrow{h}_{t+1}, \overleftarrow{C}_{t+1}), \quad (17)$$

processes it in the reverse direction. The final output is the concatenation of the two hidden states:

$$h_t = \text{concat}(\vec{h}_t, \overleftarrow{h}_t), \quad (18)$$

allowing the Bi-LSTM to integrate past and future contexts at each step [70–72].

**Remark 7.** *Bi-LSTM improves upon the unidirectional LSTM by processing data in both directions, enabling the better capture of long-term dependencies and greater prediction accuracy. This is particularly useful in natural language processing, where context depends on both preceding and succeeding terms, and in time series forecastings—such as wind speed or solar irradiance estimation—where future values can influence decisions based on past data.*

#### 4.3.2. Extreme Learning Machine (ELM)

ELM (extreme learning machine) is a feedforward neural network with a single hidden layer (SLFN, Single-Layer Feedforward Network), widely used for classification, regression, and time series forecasting due to its excellent performance [73–75]. Compared to other machine learning techniques, it offers faster learning and good performance in forecasting systems. Unlike traditional ANNs, the weights between the input and hidden layers are assigned randomly and not updated during training. The only optimized parameter is the weight between the hidden and output layers, obtained by solving a linear system [73–75]. Its architecture includes an input layer, a hidden layer with  $M$  neurons, and an output layer. Formally, if  $(X, Y) = \{(x_i, y_i)\}_{i=1}^N$  is a training dataset with  $N$  samples, where  $x_i \in \mathbb{R}^d$  is the

input vector with  $d$  features,  $y_i \in \mathbb{R}^c$  is the output vector with  $c$  classes (for classification) or real values (for regression), the output of the ELM network is expressed as  $H\beta = Y$ , where  $H$  is the activation matrix of the hidden layer,  $\beta$  is the weight matrix between the hidden layer and the output layer, and  $Y$  is the target matrix [73–75].

To compute  $H$ , if the network contains  $M$  neurons, and the activation function  $g(\cdot)$  is applied to the inputs multiplied by the random weights,  $H$  is defined as [73–75]

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_M \cdot x_1 + b_M) \\ g(w_1 \cdot x_2 + b_1) & \dots & g(w_M \cdot x_2 + b_M) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_M \cdot x_N + b_M) \end{bmatrix}_{N \times M} \quad (19)$$

where  $w_j \in \mathbb{R}^d$  is the weight vector of the  $j$ -th neuron, chosen randomly;  $b_j \in \mathbb{R}$  is the bias of the  $j$ -th neuron, also random; and  $g(\cdot)$  is the activation function (e.g., sigmoid, hyperbolic tangent, or ReLU). The determination of  $\beta$  is ensured through  $\beta = H^\dagger Y$ , where  $H^\dagger$  is the Moore–Penrose pseudoinverse of  $H$ . Obviously, if  $H$  is square and invertible,  $\beta = H^{-1}Y$ . If  $H$  is not square, the solution can be calculated using the Moore–Penrose pseudoinverse, which minimizes the least-squares term,  $H^\dagger = (H^T H)^{-1} H^T$ , from which  $\beta = (H^T H)^{-1} H^T Y$ . Once the weights  $\beta$  are obtained, the network output for a new input  $x$  is calculated as  $\hat{y} = g(Wx + b)\beta$ , where  $W$  is the matrix of random weights between the input and the hidden layer, and  $B$  is the bias vector [73–75].

**Remark 8.** ELM is characterized by its fast learning speed, thanks to randomly assigned weights between the input and hidden layers that do not require iterative updates. It offers good generalization on large datasets and employs a closed-form solution, avoiding gradient computation. These features make it effective for regression, classification, and time series forecasting tasks.

#### 4.3.3. Support Vector Regression

SVR is an extension of SVM for regression, based on nonlinear kernel functions and support vectors to build accurate predictive models. The algorithm searches for a linear regression function in a high-dimensional space, solving a quadratic programming problem [76–79]. Its accuracy and stability depend on the choice of parameters (Parameter Tuning, PT) and the selection of features (Feature Selection, FS) [35].

SVR is based on finding a regression function that minimizes the prediction error while respecting a tolerance  $\epsilon$ , which is an interval within which the error is not considered significant. The predictive function of SVR is expressed as  $f(x) = \sum_{i=1}^m \beta_i K(x, x_i) + b$ , where  $x$  is the input data,  $m$  is the number of support vectors,  $b$  is the bias term,  $\beta_i$  are the Lagrange coefficients associated with each support vector obtained during the training phase, and  $K(x, x_i)$  is the kernel function that measures the similarity between the input  $x$  and the support vector  $x_i$ . Commonly used kernel functions include linear kernels,  $K(x, x_i) = x \cdot x_i$ , polynomial kernels,  $K(x, x_i) = (x \cdot x_i + c)^d$ , radial basis function (RBF) kernels,  $K(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$ , and sigmoid kernels,  $K(x, x_i) = \tanh(\gamma x \cdot x_i + c)$  [35]. To minimize the difference between predictions and real values, but with a tolerance  $\epsilon$ , SVR involves the formulation of a loss function,  $\epsilon$ -Insensitive, defined as [35]

$$L_\epsilon(\hat{y}, y) = \begin{cases} 0, & \text{se } |\hat{y} - y| \leq \epsilon \\ |\hat{y} - y| - \epsilon, & \text{otherwise} \end{cases} \quad (20)$$

This establishes that only errors greater than  $\epsilon$  contribute to the penalization. In (20),  $\hat{y}$  is the predicted value while  $y$  is the real value, minimizing the sum of errors while maintaining the error margin defined by  $\epsilon$  [35].

SVR seeks a linear regression function,  $f(x) = w^T x + b$ , where  $w$  is the weight vector and, as usual,  $b$  is the bias. Then, by introducing a regularization parameter balancing the trade-off between error and model complexity,  $C$ , and considering the slack variables,  $\delta_i, \delta_i^*$ , which measure the error beyond the margin  $\epsilon$ , the problem becomes finding  $w$  and  $b$  values that minimize the cost function  $\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\delta_i + \delta_i^*)$ , subject to the constraints  $y_i - w^T x_i - b \leq \epsilon + \delta_i, w^T x_i + b - y_i \leq \epsilon + \delta_i^*$ , and  $\delta_i, \delta_i^* \geq 0$ , where  $\sum_{i=1}^N (\delta_i + \delta_i^*)$  represents the sum of the errors beyond the  $\epsilon$  threshold, which is penalized in the cost function [35].

**Remark 9.** *The  $\epsilon$ -insensitive function selects which errors to penalize, while the cost function quantifies and minimizes them along with the model's complexity. By defining a threshold  $\epsilon$ , the loss function ignores minor errors, while the cost function uses slack variables  $\delta_i, \delta_i^*$  to penalize only those exceeding the threshold. This balance allows SVR to trade off simplicity and accuracy, enhancing generalization and reducing sensitivity to small errors.*

During training, the SVR selects the Support Vectors, the points closest to the regression function, thus determining its shape. They are used to calculate the separating hyperplane, optimizing the error margin  $\epsilon$ .

#### 4.4. Hybrid Strategy Based in Least-Squares Regression (LSR)

The hybrid strategy based on LSR combines multiple predictive components into a single linear model, optimizing the coefficients by minimizing the sum of squared errors (SSEs) [80–82]. Formally, if  $y$  is the observed time series and  $\hat{y}$  is the estimated series as a linear combination of multiple predictors, the hybrid model takes the form  $\hat{y} = Xw + b$ , where  $X \in \mathbb{R}^{N \times M}$  is the predictor matrix, with  $N$  samples and  $M$  predictors, and, as usual,  $w \in \mathbb{R}^M$  is the vector of regression coefficients to be estimated, while  $b \in \mathbb{R}$  is the bias term. Then, starting from  $SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$ , it makes sense to write  $SSE = \sum_{i=1}^N (y_i - X_i w - b)^2$ , which, in matrix form, considering that  $\mathbf{1} \in \mathbb{R}^N$ , gives  $SSE = \|y - Xw - b\mathbf{1}\|^2$  [80–82]. Then, imposing the optimality condition,  $\frac{\partial SSE}{\partial w} = 0$ , we obtain  $w = (X^T X)^{-1} X^T y$ . Obviously, if  $X^T X$  is non-singular, we define the pseudoinverse of  $X$ ,  $X^+ = (X^T X)^{-1} X^T$ , which provides the solution  $w = X^+ y$ . Furthermore, by including the bias  $b$ , it makes sense to write  $\tilde{X} = \begin{bmatrix} X & \mathbf{1} \end{bmatrix}$  and  $w' = \begin{bmatrix} w \\ b \end{bmatrix}$ , which gives the general solution  $w' = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$  [80–82].

The LSR hybrid strategy combines multiple predictors into a single estimate, optimizing the weights to minimize the squared error. The pseudoinverse allows for a solution in both overdetermined ( $N > M$ ) and underdetermined ( $N < M$ ) systems. However, when the predictors  $X$  are correlated, the problem of multicollinearity may arise, compromising the solution's stability. A penalization  $\lambda$  is introduced to address this issue, resulting in ridge regularized least-squares regression. In this case, the optimized solution is given by  $w = (X^T X + \lambda I)^{-1} X^T y$ , where  $I$  is the identity matrix and  $\lambda$  is the regularization parameter, which helps to reduce the model's variance and improve generalization [80–82].

## 5. Experimental Setup

In this section, we compare the performance effectiveness of the VMD-SE-DL-LSR and ICEEMD-SE-DL-LSR models with those of several benchmark models. To this end, we use two distinct data sources, based on global horizontal irradiance (GHI) and wind speed (WS), collected from three different locations. This approach allows us to accurately assess the predictive capabilities of the proposed models in varying environmental and meteorological contexts, providing a thorough evaluation of their robustness and reliability compared to other reference models.

### 5.1. Time Series Datasets

This study uses time series analysis to estimate the expected future values by leveraging the already known historical data. To this end, the forecasting model is formulated as a mathematical function that estimates the value of the variable  $P$  at time  $t + 1$  based on its past values. The model can be expressed as  $P_{t+1} = \tilde{f}(y_t, y_{t-1}, \dots, y_{t-M-1})$ , where  $P_{t+1}$  represents the predicted value of the variable of interest at time  $t + 1$ ;  $y_t, y_{t-1}, \dots, y_{t-M-1}$  indicate the observed data at times  $t, t - 1, \dots, t - M - 1$  (i.e., the historical values of the time series that are used as inputs for the forecasting model). Finally,  $M$  represents the number of past values considered for making the forecast, determining the depth of the model's memory.

In our study, we adopt a short-term forecasting approach, where the model estimates a single step ahead to provide timely estimates and support quick decision-making. To improve its accuracy, we use two sources of renewable energy, which refine the model's learning and increase its ability to capture variations in the historical data, thereby optimizing the quality of the predictions.

The sampling experimental data were divided into two sections: a training section corresponding to 70% of the samples and the remaining part (30%) as a test database.

#### 5.1.1. Solar Radiation Data Set (SR)

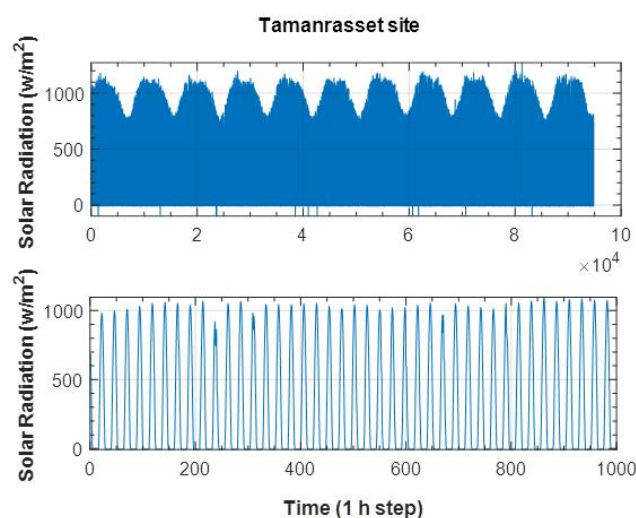
The first dataset analyzed in this study consists of eleven years of hourly observations of global horizontal irradiance, commonly called GHI, collected at the Tamanrasset station in Algeria. This station is part of the global Baseline Surface Radiation Network (BSRN) and is accessible through the portal <http://bsrn.awi.de/> (accessed on 1 May 2025). The second dataset is a one-year time series obtained through five-minute temporal resolution recordings of global horizontal irradiance at a location in Brasilia characterized by a tropical climate. These measurements were made with the National Renewable Energy Laboratory (NREL) and are available at <http://redc.nrel.gov/solar/newdata> (accessed on 1 May 2025). To ensure rigorous analysis, both datasets were collected and evaluated according to established best practices, ensuring the quality and reliability of the measurements. A summary of the main characteristics of the two data collection locations is presented in Table 2, where data from the Colorado station (cataloged site number 1322 of 39 km<sup>2</sup> at an altitude of 1299 m, longitude  $-102.52$  degrees, and latitude 37.97 degrees north) are also included. In the modeling process, each dataset is split into two subsets, one for the training and the other for the testing phases. The training phase estimates the model parameters and defines its optimal architecture, utilizing the available historical data. Subsequently, the model's performance is independently evaluated on previously unused data to assess its generalization ability and the robustness of the predictions.

#### 5.1.2. Wind Speed Dataset (WS)

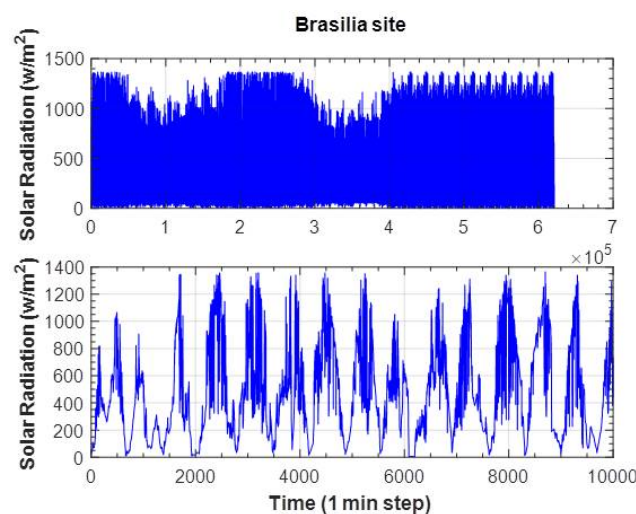
The data we have at our disposal span a decade, from 2006 to 2015. During our experiments, we worked with a total of a set of data samples with a 10-minute interval between each sample. The variation of the two datasets, solar and wind speed, are shown in Table 2. The variation of global horizontal irradiance (GHI) is illustrated for the two locations at their native resolution in Figure 2 and Figure 3, respectively. Figure 4 shows the wind speed data for the Colorado site.

**Table 2.** Some useful statistical parameters of the experimental datasets.

Site	Location	Horizon	Time Interval
1	Tamanrasset	1 h	1/1/00–31/12/10
<b>max</b> 1274	<b>min</b> 1	<b>SD</b> 340.6459	<b>Statistic information</b> <b>Mean</b> 523.4112
2	Brasilia	1 min	1/1/06–31/12/07
<b>max</b> 1368	<b>min</b> 0	<b>SD</b> 332.000	<b>Statistic information</b> <b>Mean</b> 475.5991
3	Colorado	10 min	1/1/06–31/12/15
<b>max</b> 414.60	<b>min</b> 0.100	<b>SD</b> 132.1903	<b>Statistic information</b> <b>Mean</b> 161.7558



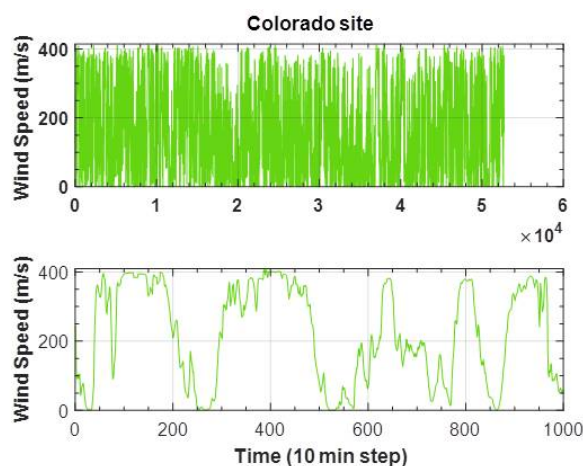
**Figure 2.** Time series of global horizontal irradiance (GHI) and wind speed (WS) at the Tamanrasset site. The upper plot shows hourly GHI data over a two-year period, while the lower plot displays wind speed measurements with a one-minute resolution. These signals serve as the input data for the forecasting framework.



**Figure 3.** Time series data for the Brasilia site: the upper panel shows the global horizontal irradiance (GHI) over an eleven-year period, while the lower panel illustrates the wind speed (WS) recorded with an hourly time step. These signals serve as the input data for the forecasting models evaluated in this study.

As an initial step, each data series is divided into two distinct subsets, serving training and testing purposes. Model parameter estimation and the optimal model architecture are determined based on the training set. Finally, the model's performance on previously unseen data is independently assessed using the testing set.

Also, in this case, 70% of the experimental data were used as a training database, while the remaining 30% were used as a testing database.



**Figure 4.** Wind speed data for the Colorado site. The upper panel shows the time series over a ten-year period, while the lower panel presents the wind speed measurements sampled at ten-minute intervals. These data serve as the basis for the forecasting experiments conducted in this study.

## 5.2. SR and WS Normalization and Evaluation Criteria

### Z-Score Normalization

Z-score normalization is a statistical technique used to transform data into a standardized scale, making the different features of a dataset comparable. This transformation is particularly useful in machine learning models as it ensures that all variables have a distribution with a mean of zero and a unit standard deviation, preventing features with different scales from disproportionately influencing the learning process.

Formally, it transforms each value  $y_i$  of a variable  $Y$  into its corresponding standardized version  $z_i$  using

$$z_i = \frac{y_i - \mu}{\sigma} \quad (21)$$

where  $z_i$  is the normalized value of the variable  $y_i$ ,  $\mu$ , as usual, is the arithmetic mean of the variable, and  $\sigma$  is its standard deviation. The Z-score transformation converts the data into standard deviation units from the mean, indicating how many standard deviations a value  $y_i$  deviates from  $\mu$ . Once normalization is performed, the mean of the transformed data will always be zero, while the standard deviation will be unitary. Naturally, values above the mean will have  $z_i > 0$ , while those below will be characterized by  $z_i < 0$ . Z-score normalization improves the convergence of machine learning algorithms, avoids scale issues between variables, facilitates data interpretation, and optimizes the performance of distance-based models, ensuring more accurate and reliable analyses. It is worth noting that normalization is not always necessary, particularly when working with data unseen by the trained model. In some cases, using raw data to preserve their interpretability and original structure may be more advantageous.

To assess the predictive performance, we used four statistical metrics: Mean Absolute Percentage Error (MAPE), Normalized Mean Squared Error (NMSE), Root Mean Squared Error (RMSE), and coefficient of determination ( $R^2$ ).

Let  $N$  be the number of testing samples,  $P_i^{\text{forecasted}}$  the forecasted value, and  $P_i^{\text{real}}$  the real value. The MAE measures the mean absolute error between the real values  $P_i^{\text{real}}$  and the predicted values  $P_i^{\text{forecasted}}$ .

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |P_i^{\text{real}} - P_i^{\text{forecasted}}|, \quad (22)$$

providing a direct measure of the mean absolute error, indicating how much, on average, the prediction differs from the real value. Additionally, the RMSE, defined as the square root of the mean squared error,

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i^{\text{real}} - P_i^{\text{forecasted}})^2} \quad (23)$$

penalizes larger errors more than the MAE, providing a more stringent measure of the prediction quality. Additionally, the NRMSE, a normalized version of the RMSE, allows for comparison between datasets with different scales. While it can be formulated in different ways, in this paper, we use the following formulation:

$$\text{NRMSE} = \frac{\text{RMSE}}{P_{\max}^{\text{real}} - P_{\min}^{\text{real}}} \quad (24)$$

where  $P_{\max}^{\text{real}}$  and  $P_{\min}^{\text{real}}$  are, respectively, the maximum and minimum of the observed values, allowing the error to be evaluated in a relative manner, independent of the data's unit of measurement. Finally,  $R^2$ , measuring how well the model explains the variability of the real data, can be formulated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (P_i^{\text{real}} - P_i^{\text{forecasted}})^2}{\sum_{i=1}^n (P_i^{\text{real}} - \bar{P}^{\text{real}})^2} \quad (25)$$

where  $\sum_{i=1}^n (P_i^{\text{real}} - P_i^{\text{forecasted}})^2$  is the sum of squared residuals, while  $\sum_{i=1}^n (P_i^{\text{real}} - \bar{P}^{\text{real}})^2$  is the total sum of squares ( $R^2$  close to 1 indicates that the model is highly predictive).

### 5.3. Prediction Models for SR and WS: Comparative Procedure

This study uses two different types of models for comparison: single and hybrid models based on the VMD-SE and ICEEMDAN-SE decomposition techniques. The accuracy of these methods is of considerable importance for predicting wind speed (WSP) and solar radiation (SRP), making it necessary to evaluate their effectiveness and performance rigorously.

In the first phase of the procedure, VMD and ICEEMDAN are used to decompose the datasets related to WS and SR into distinct sub-signals, the details of which are provided in Table 3. From this, it is evident that VMD uses a bandwidth constraint of 2000 to ensure optimal separation between the signal's different modes while the number of extracted modes is set to 12. On the other hand, ICEEMDAN, based on a standard deviation control of white noise, set to 0.2, performs one hundred iterations to improve the decomposition stability. These parameters directly affect the quality of the signal segmentation and, consequently, the model's ability to identify the most relevant features for forecasting.

Once the sub-signals are obtained, the most significant ones are selected based on the sample SE value used in the forecasting phase. To improve the accuracy of the estimates, the model combines the results obtained using BiLSTM, ELM, and SVM, allowing for the integration of information extracted from each method. Finally, the final prediction is obtained by applying a combined strategy based on LSR, yielding a more accurate prediction than using a single model, thereby optimizing the system's overall performance.

**Remark 10.** *The selection of the number of modes in VMD and ICEEMDAN is a very important task, especially considering that there is no universally accepted rule for determining this parameter. In our study, we conducted a series of exploratory trials to identify the configuration that most accurately forecasted our desired results. After careful examination, we concluded that, for VMD, using 12 modes, imposing a bandwidth constraint of 2000, and performing 100 iterations with  $\varepsilon = 0.2$  in ICEEMDAN yielded the most accurate and stable outcomes. Although the model was applied to different locations with notably different weather conditions, we chose to maintain the same parameter values across all sites rather than adjusting them individually, as this approach performed better in our assessments. Despite the fact that the technique is applied on different sites, the parameter values remain the same, since these are the best choices and give the best results.*

**Table 3.** Configuration of the decomposition procedures applied in the proposed forecasting framework. (a) ICEEMDAN employs white noise with a standard deviation of 0.2 and 100 iterations to enhance the stability and accuracy of the intrinsic mode function (IMF) separation; (b) VMD applies a bandwidth constraint of 2000 and decomposes the signal into 12 distinct modes, improving feature extraction for global horizontal irradiance (GHI) and wind speed (WS) forecasting.

Method	Parameter	Value
ICEEMDAN	Standard deviation of the white noise $\varepsilon$	0.2
	Number of trials $I$	100
VMD	Bandwidth constraint $\alpha$	2000
	Number of modes $K$	12

## 6. Computational Complexity Analysis and Model Efficiency Evaluation

The implementation of the proposed framework requires a thorough evaluation of computational complexity, as the forecasting process consists of several stages, each with a significant impact on processing time. The main components of the approach include signal decomposition, entropy-based selection, the training of predictive models, and forecast hybrid using least-squares regression (LSR).

Signal decomposition is one of the most computationally intensive phases. VMD, being an iterative method that optimizes the energy of modal components, has a complexity of  $O(KN \log N)$ , where  $K$  is the number of extracted modes and  $N$  is the length of the time series. In experimental tests conducted on datasets with different temporal resolutions, the average computation time for VMD decomposition was approximately 3.2 s for 10,000 samples with 12 extracted modes.

ICEEMDAN, while offering a more refined separation of modal components, introduces a higher computational cost, estimated at  $O(NIK)$ , where  $I$  represents the number of iterations required to stabilize the process. Experimental analysis strongly suggested that processing ICEEMDAN on a 10,000-sample dataset required an average of 8.7 s, making it significantly more expensive than VMD.

The entropy-based selection phase, using SE, adds another layer of complexity, with a computational cost of  $O(N^2)$  when applied to all extracted modes. A reduced selection strategy was implemented to improve efficiency, analyzing only the modes with the highest information content and reducing the average computation time to 2.1 s for 10,000 samples.

Training the predictive models varies in complexity depending on the technique used. Bi-LSTM, which leverages a recurrent neural network, has a cost of  $O(NH^2)$ , where  $H$  is the number of hidden units in the network. Training a Bi-LSTM model on a 10,000-sample dataset in experimental tests required approximately 12.5 s, highlighting its high computational cost.

ELM, due to its simplified architecture, allows for high-speed training with a complexity of  $O(NM)$ , where  $M$  is the number of neurons in the hidden layer. In the tests

conducted, ELM showed an average training time of less than 0.8 s, making it the most efficient algorithm among those considered.

Conversely, SVR has a computational complexity of  $O(N^2)$  for nonlinear kernels, making it less scalable for large datasets. Execution time analysis indicated that training SVR on 10,000 samples required around 6.3 s, placing it between Bi-LSTM and ELM in terms of efficiency.

LSR was evaluated as computationally efficient, with a complexity of  $O(NM^2)$ , where  $M$  is the number of aggregated models. The average computation time for this phase was 0.5 s, making it negligible compared to the other stages.

The comparative analysis of execution times indicated that the total time required for the entire framework ranges from 18.2 to 34.9 s, depending on the decomposition technique used. The use of VMD proved to be the most advantageous in terms of efficiency, while ICEEMDAN significantly increased the computational cost without providing a proportional improvement in forecast accuracy.

The main configuration parameters and training settings for the Bi-LSTM, ELM, and SVR models are summarized in Tables 4 and 5. These results demonstrate that the developed framework ensures high accuracy with computation times compatible with operational applications, suggesting opportunities for further optimization in future real-world implementations.

**Table 4.** Model configuration parameters used for solar radiation and wind speed forecasting at the three study sites (Tamanrasset, Brasilia, and Colorado). The table reports the main settings for each predictive model—Support Vector Machine (SVM), extreme learning machine (ELM), and Bidirectional Long Short-Term Memory (Bi-LSTM)—including kernel type, regularization parameters, number of neurons and layers, activation functions, dropout rates, and optimizer type.

Models	Parameters	Tamanrasset	Colorado	Brazilia
SVM	Kernel	RBF	RBF	RBF
	C	11	6	11
	$\gamma$	220	120	223
ELM	neurons	120	50	48
	activation	sigmoid	sigmoid	sigmoid
BiLSTM	layers	2	1	1
	units	101	107	97
	dropout	0.0021	0.000021	0.000021
	optimizer	adam	adam	adam

**Table 5.** Average training time (in seconds) required by each predictive model—SVM, ELM, Bi-LSTM, and LSR—across the three experimental datasets (Tamanrasset, Colorado, and Brasilia). These results highlight the computational efficiency of each model within the proposed hybrid forecasting framework.

Models/Time (s)	Tamanrasset	Colorado	Brazilia
SVM	5	2	8
ELM	3	1	5
BiLSTM	98	158	100
LSR	$\leq 0.5$	$\leq 0.5$	$\leq 0.5$

## 7. Experimental Results

### 7.1. Results of Decomposition Step

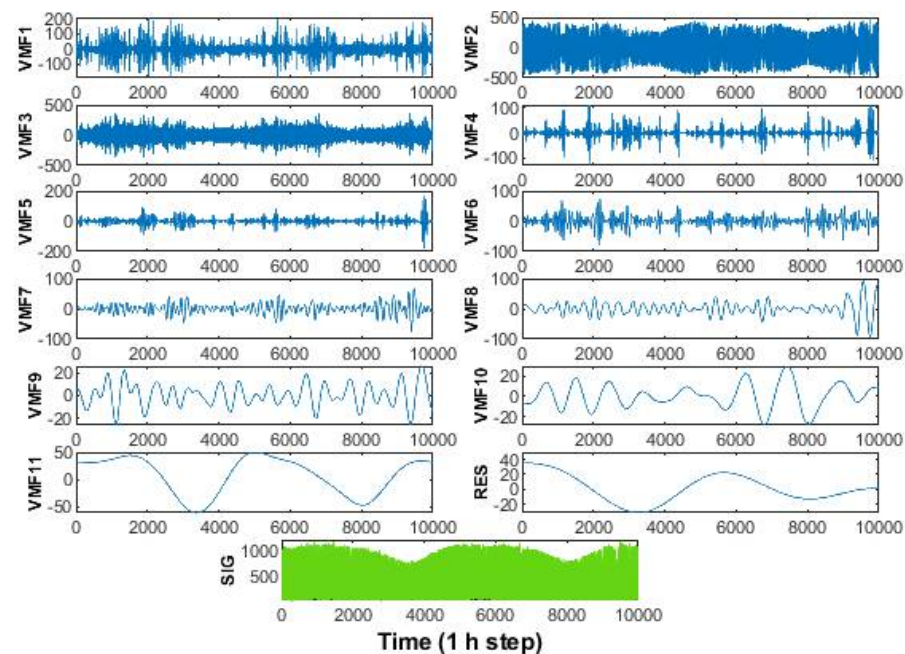
#### 7.1.1. Decomposition by VMD

VMD has proven effective in decomposing complex energy signals, significantly improving the forecasting of global solar radiation and wind power (SD). The analysis of the results highlighted that the optimal selection of decomposition parameters directly affects

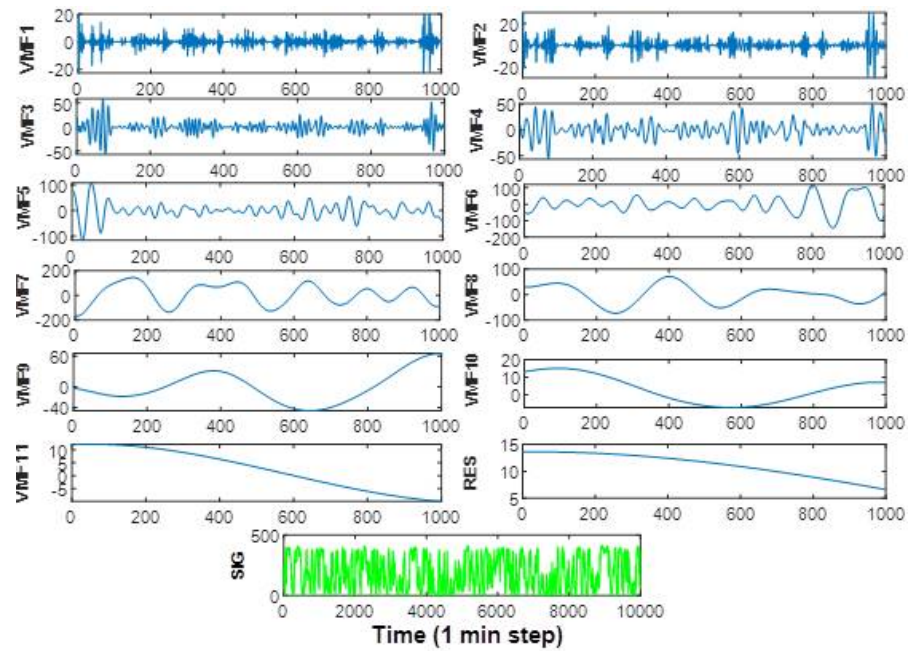
the predictive model's performance, strongly impacting the quality of mode separation and the ability to capture the fundamental characteristics of the signal. One of the most relevant findings was the bandwidth constraint's role in determining the decomposition quality. Experimental data showed that setting this parameter to 200 allowed for a clear separation of modal components without frequency overlaps. Comparative analysis with other methods revealed that a value too low would result in overly broad modes, potentially merging multiple frequencies, while a higher value would increase computational complexity without significant gains in predictive accuracy.

Another key element was the Lagrangian stop multiplier, which was set as zero divided by the parameter  $r$ . This configuration avoided additional penalties on mode deviation during iterations, allowing the model to adapt more naturally to the signal structure. The convergence of the iterative process was determined solely by the predefined tolerance level, ensuring a balance between accuracy and stability in the decomposition. The method also excluded the zero-frequency component, removing any constant mean value from the signal. This made it possible to focus exclusively on dynamic variations, enhancing the ability to identify meaningful patterns for forecasting. Central frequencies of the modes were uniformly initialized, ensuring a balanced distribution of components across the frequency spectrum.

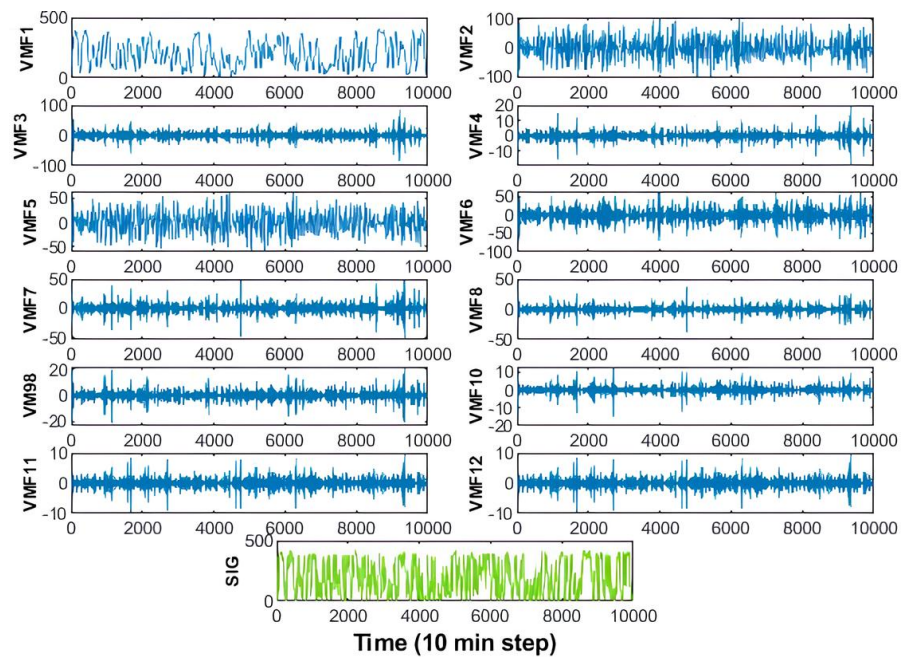
The decomposition using VMD was applied to time series solar radiation and wind power data collected from the Tamanrasset, Brasilia, and Colorado sites. As shown in Figures 5–7, the signals were divided into ten main modes plus one residual. This decomposition enabled the precise isolation of the different frequency components, simplifying the modeling process for forecasting tasks.



**Figure 5.** Variational mode decomposition (VMD) applied to the Tamanrasset dataset. The original signals of global horizontal irradiance (GHI) and wind speed (WS) are decomposed into a set of intrinsic mode functions (IMFs), allowing the isolation of meaningful frequency components. This process reduces background noise and enhances the quality of input data for forecasting, thereby improving the accuracy of the predictive models.



**Figure 6.** Application of variational mode decomposition (VMD) to the dataset from the Brasília site. The original time series of global horizontal irradiance (GHI) and wind speed (WS) is decomposed into intrinsic mode functions (IMFs). This decomposition isolates key frequency components, reduces background noise, and enhances the interpretability and predictive accuracy of the forecasting models.



**Figure 7.** Application of variational mode decomposition (VMD) to the Colorado site dataset. The original global horizontal irradiance (GHI) signal is decomposed into a set of intrinsic mode functions (IMFs), allowing the extraction of key frequency components. This decomposition reduces noise and enhances the identification of meaningful patterns, thereby improving the accuracy of solar radiation forecasting.

The analysis of sample entropy (SE) applied to the extracted IMFs revealed that the highest values were observed in modes five to ten for the Tamanrasset and Brasilia sites, whereas in the Colorado site, the highest entropy values were concentrated in the first three modes. Specifically, the maximum entropy value for Tamanrasset reached 1.650 and 1.514 for Brasilia. In contrast, the values for Colorado were lower, with a maximum of 1.284 in

the first mode. This suggests that in the first two locations, the higher modes contain more informational complexity, while in Colorado, the most relevant information for prediction is concentrated in the initial components.

The method's effectiveness was further confirmed by the results obtained in solar radiation and wind speed forecasting. Data from the Tamanrasset site showed that the VMD-based model, combined with sample entropy, artificial neural networks (ANNs), and least-squares regression, achieved a root mean square error (RMSE) of 9.9046 and a mean absolute error (MAE) of 0.0137. The coefficient of determination reached 0.9990, the highest among all models analyzed. For the Brasilia site, slightly lower values were recorded, with an RMSE of 10.4583 and a coefficient of determination of 0.9954. The Colorado site showed slightly lower performance than the other two, with an RMSE of 12.7856 and a coefficient of determination of 0.9921, though still outperforming the competing models tested in the study.

Compared to other approaches, the hybrid VMD-based model significantly improved predictive accuracy over individual ML techniques such as neural networks or support vector regression, demonstrating greater adaptability to the non-stationary nature of energy data. This result was especially evident at the Tamanrasset site, where solar radiation variability was high, and the decomposition strategy effectively isolated the most impactful fluctuations.

**Remark 11.** *It is worth noting that  $\tau$  is not the same as  $\lambda(t)$  in Equation (2). In fact,  $\lambda(t)$  is used to enforce the reconstruction constraint of the original signal from the extracted modes in the VMD optimization problem. This parameter is updated iteratively within the Lagrange multipliers method to ensure that the sum of the modes reconstructs the original signal. On the other hand,  $\tau$  is the stopping criterion that determines when the VMD iterative process should terminate. It defines the convergence threshold, i.e., the level of variation between two successive iterations below which the decomposition is considered stable and complete.*

#### 7.1.2. Decomposition by ICEEMDAN

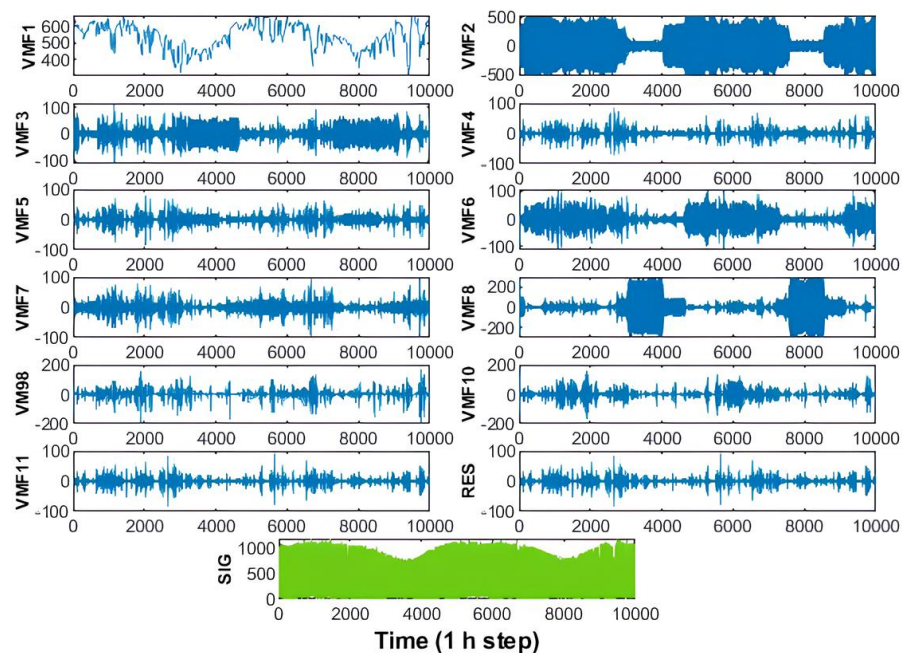
The analysis of the results obtained using the ICEEMDAN method highlights how this technique enabled a detailed and accurate decomposition of RES signals from the three study sites, as illustrated in Figures 8–10. Compared to the VMD method, which decomposed the signal into a fixed number of modes with predefined frequency constraints (Figures 5–7), ICEEMDAN showed greater flexibility in separating components, better adapting to the specific characteristics of each dataset.

One of ICEEMDAN's main advantages is its ability to reduce mode mixing, ensuring a clearer separation of the intrinsic mode functions. The decomposition generated a series of modes, each associated with a specific frequency range, allowing for the more effective extraction of features useful for forecasting global solar radiation and wind speed.

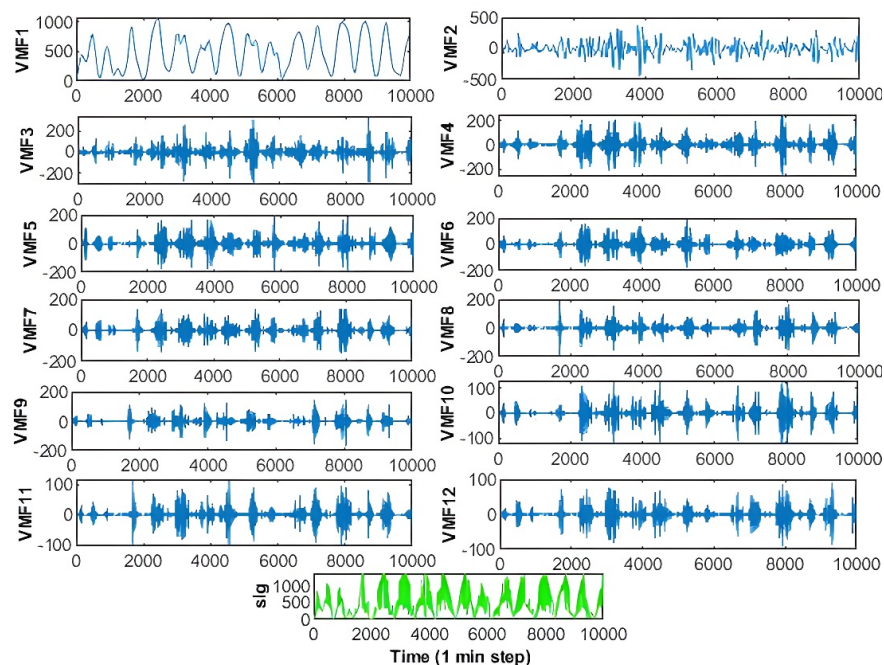
For the Tamanrasset site, ICEEMDAN decomposition showed that the most informative modes were mainly concentrated between the fifth and tenth modes. The sample entropy (SE) calculated for each mode reached a maximum value of 1.650, indicating significant complexity in the higher-order informative components.

As shown in Figure 8, the initial modes were dominated by high-frequency fluctuations, mainly associated with noise and short-term variations, while the subsequent modes contained more structured signals with relevant predictive information. The final residual captured the overall trend of the signal, helping to identify long-term patterns. Comparing these results to those obtained with VMD (see Figure 5), a clear difference emerges in how information is distributed across extracted modes. While VMD divides the signal into a predefined number of modes with fixed frequency bands, ICEEMDAN produces modes

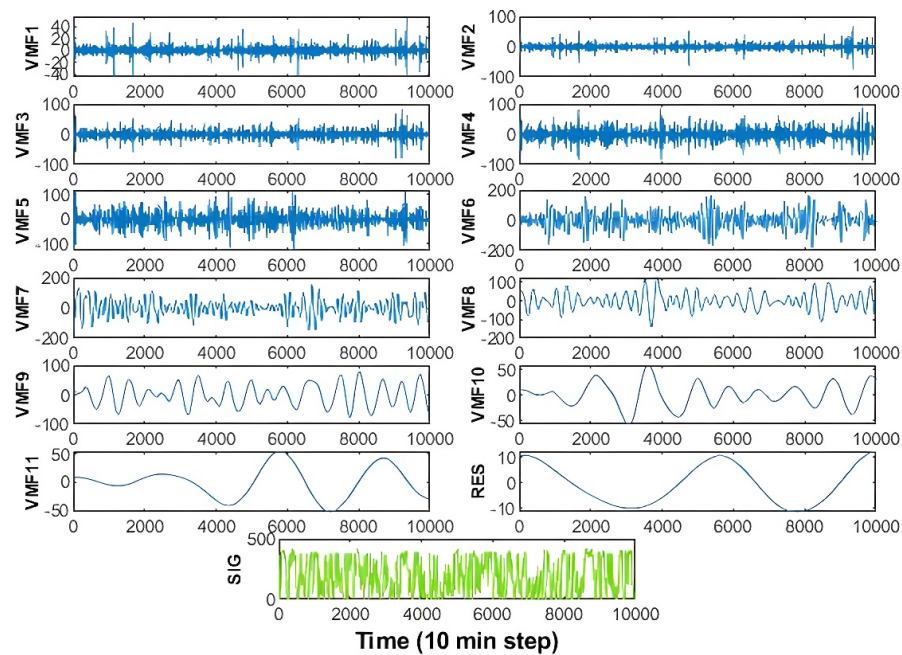
that dynamically adapt to the signal's structure. This leads to more effective noise reduction and clearer separation of meaningful components.



**Figure 8.** ICEEMDAN-based signal decomposition for the Tamanrasset site. The original signal is decomposed into intrinsic mode functions (IMFs), allowing the isolation of key frequency components while reducing background noise. This preprocessing step enhances the quality of the input data and contributes to improved forecasting performance for global horizontal irradiance (GHI) and wind speed (WS).



**Figure 9.** ICEEMDAN-based signal decomposition for the Brasilia site. The original time series is decomposed into intrinsic mode functions (IMFs), allowing the isolation of key frequency components while reducing noise. This process enhances the quality of the global horizontal irradiance (GHI) and wind speed (WS) signals, thereby improving the accuracy of the subsequent forecasting models.



**Figure 10.** Application of ICEEMDAN to the Colorado dataset. The original wind speed signal is decomposed into a series of intrinsic mode functions (IMFs), enabling the extraction of significant frequency components. This decomposition reduces noise and enhances the predictive accuracy of the forecasting models.

ICEEMDAN produced a different information distribution for the Brasilia site compared to Tamanrasset. SE recorded a maximum value of 1.514, slightly lower than in the first case, indicating lower complexity in the higher-order modes. The modal analysis in Figure 9 revealed that intermediate modes, particularly between IMF4 and IMF7, contained most of the relevant predictive information. In contrast, the first modes had significant noise, while the residual showed a regular pattern with evident seasonal cycles. A comparison with VMD decomposition (Figure 6) confirmed that ICEEMDAN ensured more effective separation between informative and noisy components. With VMD, some modes overlapped more regarding information content, while ICEEMDAN provided a clearer distribution across modes, improving the distinction between relevant and irrelevant signals.

ICEEMDAN revealed a different information distribution for the Colorado site from the previous two cases. SE reached a maximum value of 1.284, the lowest among the three sites analyzed, suggesting lower complexity in the informative components. As shown in Figure 10, the first three modes contained most of the significant predictive information, while the higher-order modes were less relevant than those at the other sites. The residual exhibited less pronounced fluctuations, indicating more stable meteorological conditions in Colorado. Compared to the VMD decomposition (see Figure 7), ICEEMDAN extracted modes with a more concentrated distribution of information in the initial components, while VMD provided a more uniform spread across the frequency spectrum. This implies that, for the Colorado site, ICEEMDAN was more effective in capturing essential information with fewer components, simplifying the predictive modeling process.

The results show that the two decomposition methods differ significantly in information distribution and signal component separation. ICEEMDAN demonstrated greater flexibility in mode allocation, adapting better to the specific characteristics of each site. With a fixed number of modes and predefined frequency bands, VMD produced a more uniform decomposition but showed lower adaptability to local signal variations. At the Tamanrasset and Brasilia sites, ICEEMDAN concentrated information in the higher-order modes, while in Colorado, it focused on the initial components, demonstrating its adaptability to local

conditions. ICEEMDAN more effectively reduced high-frequency noise and separated informative components with greater clarity compared to VMD.

These differences between decomposition methods directly impacted the performance of the forecasting models. The prediction results indicated that the ICEEMDAN-based model, combined with SE and ANN, achieved slightly better performance than the VMD-based model. Specifically, for the Tamanrasset site, ICEEMDAN reached a root mean square error (RMSE) of 9.7498, slightly lower than the 9.9046 obtained with VMD. For the Brasilia site, the RMSE with ICEEMDAN was 10.2157, compared to 10.4583 with VMD, confirming a slight advantage for the former. Finally, for the Colorado site, ICEEMDAN achieved an RMSE of 12.5132, outperforming the 12.7856 recorded with VMD.

These results demonstrate that combining ICEEMDAN decomposition, component selection via sample entropy, and deep learning models provides more accurate forecasts than those based on VMD. ICEEMDAN's enhanced adaptability to the specific characteristics of the signals improved the quality of information separation, reduced overall error, and increased the reliability of the predictions.

### 7.1.3. Application of SE

The analysis of SE represents a crucial step in selecting the most relevant modal components after time series decomposition using VMD and ICEEMDAN. This method makes it possible to identify the components with the highest degree of complexity and informational variability, thereby improving the quality of forecasts on renewable energy data.

As shown in Table 6, for the Tamanrasset site, the IMF1–IMF5 components derived from the ICEEMDAN decomposition exhibit significantly higher SE values compared to the remaining components, suggesting that these modes contain the most relevant information for forecasting. In particular, the SE value for IMF1 is 1.006, while for IMF5, it decreases to 0.279, indicating a progressive reduction in complexity in the lower-frequency components. Similarly, for the Brasilia site, ICEEMDAN decomposition reveals that IMF1–IMF5 holds the highest sample entropy values, with IMF1 at 0.240 and IMF5 at 0.350. This trend suggests that the data structure is characterized by greater variability in the high-frequency components. SE values show a similar distribution for the Colorado site: IMF1 records the highest value at 1.284, while IMF5 drops to 0.441. This indicates that the complexity of the information is concentrated in the higher-order components, making them essential for forecasting.

Compared with the VMD decomposition (see Table 7), it is observed that for the Tamanrasset site, the IMF5–IMF10 components show the highest SE values, with IMF5 reaching 1.514 and IMF10 at 1.135. This suggests that VMD distributes informational complexity into lower-frequency modes compared to ICEEMDAN, potentially improving forecasting stability. For the Brasilia site, the IMF5–IMF10 components obtained from VMD report SE values ranging from 0.475 to 0.484, confirming a more homogeneous information distribution than ICEEMDAN. Finally, the maximum SE value for the Colorado site is recorded in IMF10 at 1.553, suggesting that VMD tends to retain more information in lower-frequency modes.

The results indicate that the choice of decomposition technique influences the selection of the most informative modal components. ICEEMDAN tends to concentrate on high-frequency components, making it particularly effective in capturing rapid fluctuations in the data. On the other hand, VMD better distributes informational complexity across intermediate and low-frequency modes, improving the stability of long-term forecasts.

Integrating SE-based component selection with ML models enhances forecasting accuracy. The results in Tables 6 and 7 demonstrate that selecting modal components with higher SE values leads to a significant reduction in forecasting error, suggesting that an

optimal combination of decomposition and mode selection can improve the reliability of solar and wind energy prediction models.

**Table 6.** Sample entropy (SE) values associated with each intrinsic mode function (IMF) obtained through variational mode decomposition (VMD) for the Tamanrasset, Brasilia, and Colorado sites. The figure illustrates how informational complexity is distributed across the different frequency modes, with higher SE values indicating more informative components. These findings support the entropy-based selection process used to enhance the forecasting accuracy of solar radiation and wind speed.

IMFs/SE	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
Tamansasset	0.092	0.424	0.265	0.332	<b>1.514</b>	<b>1.335</b>
Brasilia	0.027	0.127	0.305	0.380	<b>0.475</b>	<b>0.482</b>
Colorado	0.0274	0.112	0.211	0.430	0.685	0.909
IMFs/SE	IMF7	IMF8	IMF9	IMF10	IMF11	RES
Tamansasset	<b>1.650</b>	<b>1.562</b>	<b>1.434</b>	<b>1.135</b>	0.675	0.293
Brasilia	<b>0.501</b>	<b>0.498</b>	<b>0.452</b>	<b>0.484</b>	0.398	0.231
Colorado	<b>1.142</b>	<b>1.283</b>	<b>1.443</b>	<b>1.553</b>	<b>1.570</b>	<b>1.707</b>

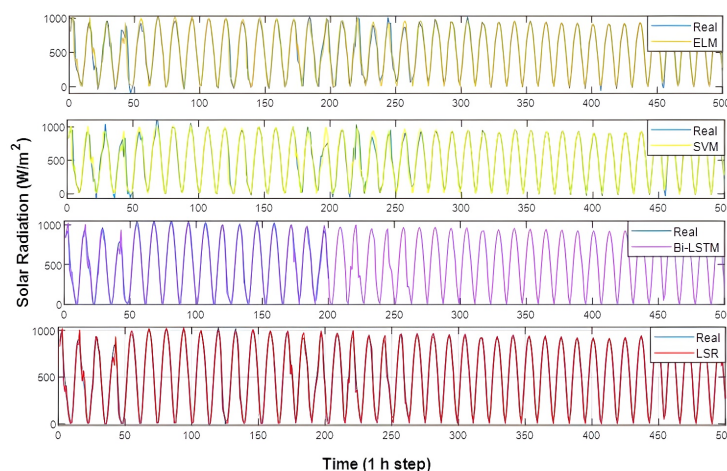
**Table 7.** Sample entropy (SE) values corresponding to each mode extracted via ICEEMDAN decomposition for the Tamanrasset, Brasilia, and Colorado sites. The chart illustrates how informational complexity is distributed across the intrinsic mode functions (IMFs), highlighting the most informative components for accurate forecasting of solar radiation and wind speed.

IMFs/SE	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
Tamansasset	<b>1.006</b>	<b>1.290</b>	<b>1.328</b>	<b>0.314</b>	0.279	0.206
Brasilia	0.240	<b>0.360</b>	<b>0.378</b>	<b>0.328</b>	<b>0.350</b>	0.159
Colorado	<b>1.284</b>	<b>1.236</b>	<b>1.021</b>	<b>0.744</b>	0.441	0.189
IMFs/SE	IMF7	IMF8	IMF9	IMF10	IMF11	RES
Tamansasset	0.102	0.042	0.032	0.014	0.003	0.002
Brasilia	0.102	0.043	0.025	0.009	0.005	0.001
Colorado	0.102	0.055	0.030	0.013	0.005	0.001

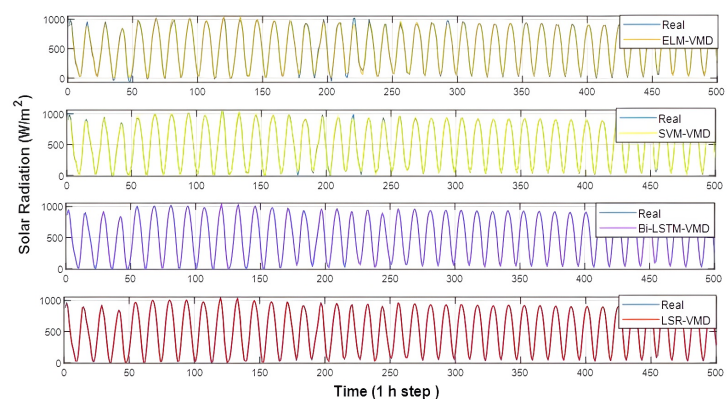
## 7.2. Performance Comparison of Forecasting Models

Without data decomposition, the predicted values from all models follow the actual values in the forecasting results but with a noticeable time lag. After decomposition using VMD and ICEEMDAN, this delay is reduced. Specifically, Figure 11 illustrates the predictive performance of the proposed model in terms of statistical accuracy, comparing the forecasts obtained through the hybrid framework with the actual observed values. The results highlight how combining decomposition techniques, sample entropy-based selection, and the hybrid of predictive models enables a highly accurate estimation of solar radiation and wind speed. In particular, a close alignment between the predicted and real data is observed, with reduced forecasting error and strong generalization capability across different geographic contexts. Moreover, Figure 12 displays the solar radiation forecasting using ELM, SVM, and Bi-LSTM (along with their combination using LSR) before decomposition for the Tamanrasset site. At the top, the original trend ( $P^{\text{real}}$ ) is shown overlaid with the forecast curve generated by ELM, while immediately below, the overlays between ( $P^{\text{real}}$ ) and the forecasted curves ( $P^{\text{forecasted}}$ ) using SVM, Bi-LSTM, and LSR, respectively, are illustrated.

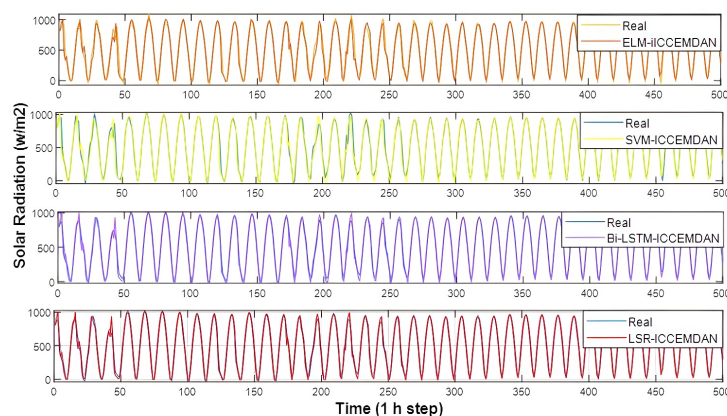
Similarly, Figures 12 and 13 show the same predictive trends for the Tamanrasset site, after decomposing the signals using VMD and ICEEMDAN.



**Figure 11.** Forecasting of solar radiation for the Tamanrasset site using three predictive models: ELM, SVM, and Bi-LSTM. The final prediction is obtained by combining the outputs of the individual models through least-squares regression (LSR).

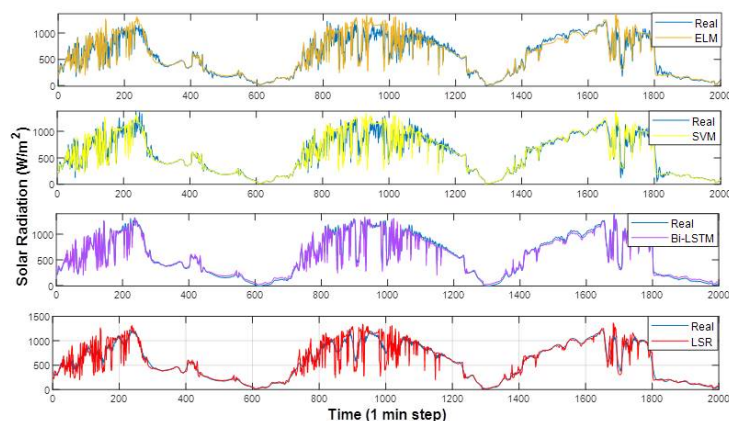


**Figure 12.** Solar radiation forecasting results for the Tamanrasset site using three predictive models—extreme learning machine (ELM), support vector regression (SVM), and Bidirectional Long Short-Term Memory (Bi-LSTM)—combined with variational mode decomposition (VMD). The outputs of the individual models are integrated using a least-squares regression (LSR) strategy to enhance prediction accuracy.

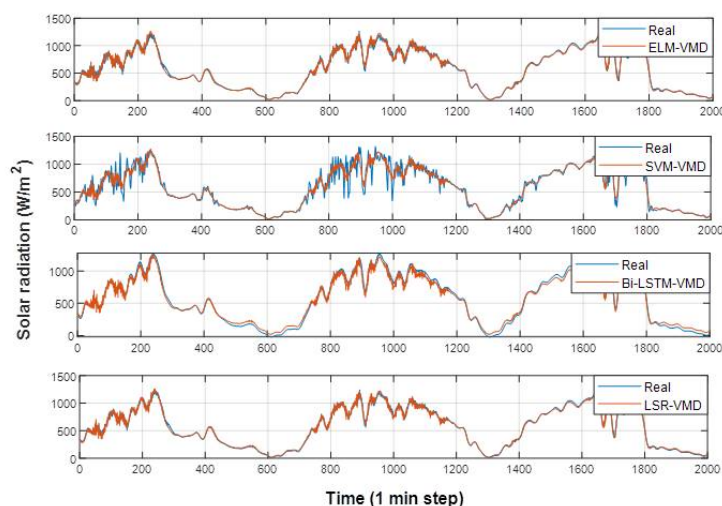


**Figure 13.** Solar radiation forecasting results for the Tamanrasset site using ELM, SVM, and Bi-LSTM models, as well as their hybrid combination through least-squares regression (LSR). The input data were preprocessed using the ICEEMDAN decomposition method to extract informative components and enhance prediction accuracy.

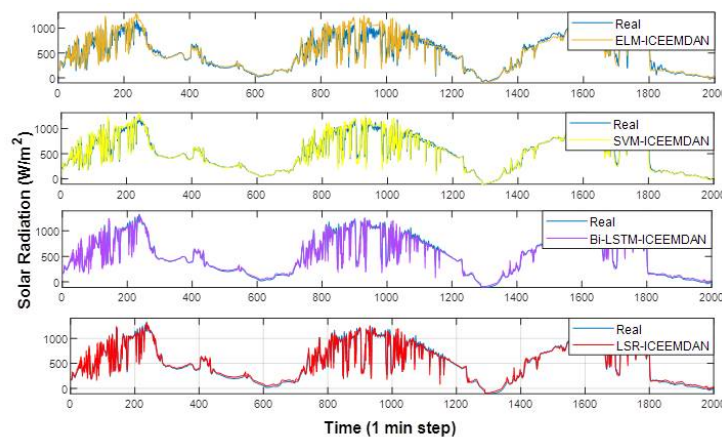
For the Brasilia site, the corresponding figures are Figures 14–16, while for the Colorado site, the forecast trends are shown in Figures 17–19.



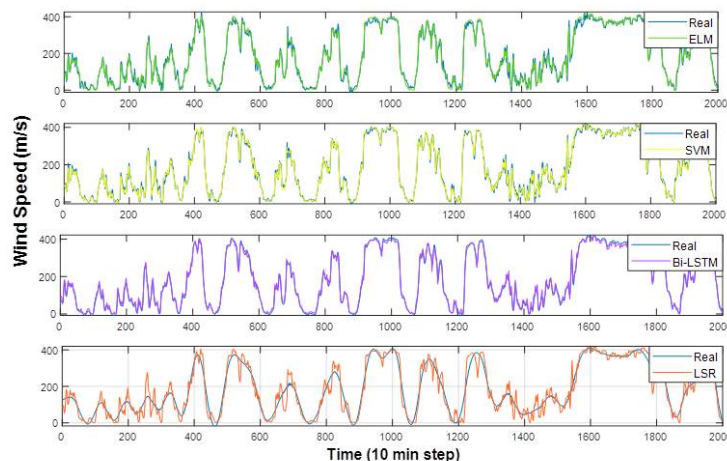
**Figure 14.** Forecasting results of solar radiation at the Brasilia site using ELM, SVM, and Bi-LSTM models applied to the raw (non-decomposed) signal. The outputs of the individual models are also combined using a least-squares regression (LSR) strategy to enhance prediction accuracy.



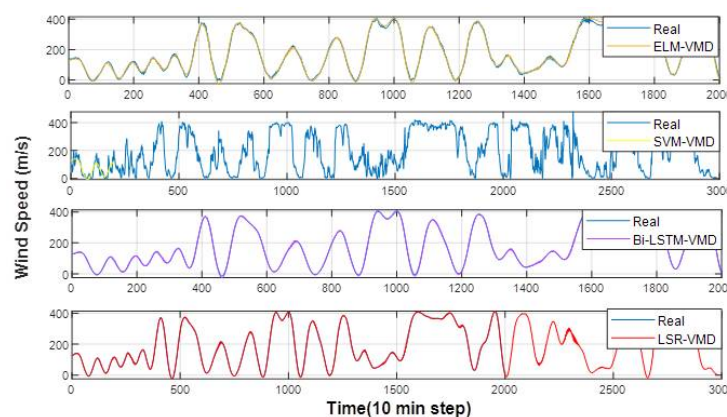
**Figure 15.** Solar radiation forecasting results for the Brasilia site using ELM, SVM, and Bi-LSTM models after signal decomposition with VMD. The outputs of the individual models are also combined through least-squares regression (LSR) to enhance prediction accuracy.



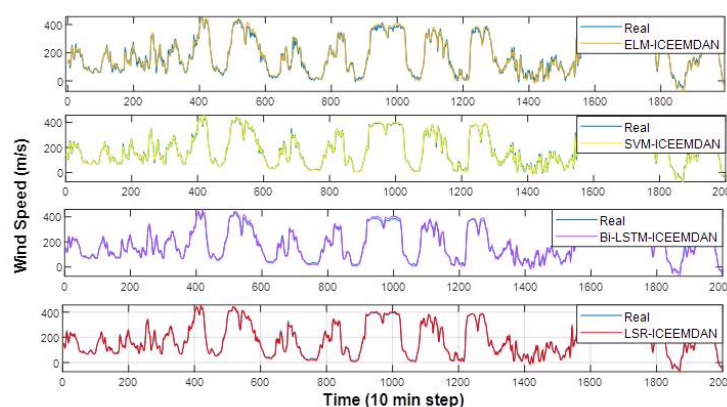
**Figure 16.** Solar radiation forecasting results for the Brasilia site using ELM, SVM, and Bi-LSTM models, combined through a least-squares regression (LSR) strategy after ICEEMDAN-based signal decomposition. The figure illustrates the predictive performance of each individual model as well as the enhanced accuracy achieved by the hybrid approach.



**Figure 17.** Wind speed forecasting results for the Colorado site using individual models (ELM, SVM, and Bi-LSTM) and their combined output via least-squares regression (LSR), applied to raw data without prior signal decomposition.



**Figure 18.** Wind speed forecasting for the Colorado site using ELM, SVM, and Bi-LSTM models after signal decomposition with VMD. The final forecast is obtained by combining the individual model outputs through least-squares regression (LSR).



**Figure 19.** Wind speed forecasting for the Colorado site using ELM, SVM, and Bi-LSTM models. The input data were first decomposed using the ICEEMDAN method, and the individual model predictions were subsequently combined using a least-squares regression (LSR) strategy to enhance accuracy.

Tables 8–10 provide a detailed overview of the weights associated with the different models for each site, namely Tamanrasset, Brasilia, and Colorado sites, respectively, and the preprocessing method.

**Table 8.** Time series of global horizontal irradiance (GHI) recorded at the Tamanrasset site (Algeria). The data cover an 11-year period and are used for solar radiation forecasting.

Weight	Bi-LSTM	SVM	ELM	Sum (wi)
without	1.1042	−0.0113	−0.1207	0.9850
VMD	0.5599	0.4932	−0.0534	0.9997
ICEEMDAN	0.5048	0.3744	0.1239	1.0031

**Table 9.** Time series of global horizontal irradiance (GHI) recorded at the Brasilia site over a one-year period with a five-minute resolution. This dataset is used to evaluate the forecasting performance of the proposed hybrid model under tropical climate conditions.

Weight	Bi-LSTM	SVM	ELM	Sum (wi)
without	1.1291	−0.0301	−0.1139	0.9851
VMD	0.7598	0.2175	0.0188	0.9961
ICEEMDAN	1.2108	−0.0439	−0.1951	0.9718

**Table 10.** Original wind speed (WS) measurements from the Colorado site, recorded at 10-minute intervals over the period 2006–2015. These data serve as the basis for evaluating the performance of the proposed hybrid forecasting framework.

Weight	Bi-LSTM	SVM	ELM	Sum (wi)
without	0.7001	0.2363	0.0590	0.9954
VMD	0.9800	0.0854	−0.0644	1.0010
ICEEMDAN	0.6500	0.4938	−0.1306	1.0131

For the Tamanrasset site, Bi-LSTM emerges as the dominant model with a weight of 1.10 without preprocessing, while SVM and ELM have very low and negative weights. When VMD is applied, the effectiveness of SVM improves significantly, with a weight reaching 49.32%, almost equivalent to Bi-LSTM's weight of 55.99%. While still providing a weak contribution, ELM shows a slight improvement with a positive weight of 0.53. With the application of ICEEMDAN, the impact on SVM is significant, with a weight of 37%, while Bi-LSTM continues to dominate with a weight of 50.48%. ELM achieves a positive weight of 12.39%, indicating that this decomposition technique allows for a more balanced contribution of all models to the prediction. Therefore, for the Tamanrasset site, ICEEMDAN seems to improve the performance of SVM and ELM, balancing the contributions of the various models.

Without any treatment, Bi-LSTM has a dominant contribution for the Brasilia site with a weight of 1.13, while SVM and ELM show negative weights of −0.03 and −0.11, respectively. The application of VMD improves the performance of all models, with BiLSTM remaining dominant with a weight of 75.98%, followed by SVM with a contribution of 21.12%, and ELM showing an improvement with a positive weight of 1.88%. With the application of ICEEMDAN, Bi-LSTM continues to dominate with a weight of 1.21, but SVM and ELM remain with inferior performances, with weights of −0.0439 and −0.1951, respectively. These results suggest that, even with preprocessing techniques such as VMD and ICEEMDAN, Bi-LSTM remains the most effective model for this site, with SVM benefiting more from VMD but never reaching the performance level of Bi-LSTM.

For the Colorado site, Bi-LSTM is again the dominant model without preprocessing, with a contribution of 70%. The SVM model shows a moderate contribution (23.6%), while ELM provides the lowest contribution (5.9%). When VMD is applied, Bi-LSTM's effectiveness increases dramatically, with a weight of 98%. ELM shows a negative weight of −6.4%, indicating that this model does not benefit from VMD preprocessing. The SVM model maintains a moderate performance with a weight of 8.5%. With the application of ICEEMDAN, the effect on Bi-LSTM is reduced, but the model remains dominant with a weight of 65%. SVM benefits from the application of ICEEMDAN, with the weight

increasing to 49.38%, while ELM continues producing poor results with a negative weight of  $-13\%$ . Overall, ICEEMDAN helps balance the contribution of the models, but Bi-LSTM remains the most effective model for the Colorado site.

Overall, the application of VMD proves to be highly beneficial for Bi-LSTM, significantly increasing its weight, especially for the Colorado site, where BiLSTM becomes the dominant model. However, VMD does not seem helpful for ELM, which sometimes shows negative weights, suggesting that ELM is unsuitable for these datasets, regardless of the preprocessing treatment. ICEEMDAN, on the other hand, proves to be particularly useful for improving the performance of SVM and, in some cases, helps balance the contribution of the various models. However, Bi-LSTM remains the dominant model in nearly all cases, with SVM showing improvements at the Brasilia and Tamanrasset sites but never surpassing Bi-LSTM's performance.

Table 11 summarizes the evaluation metric results for the Tamanrasset site using the various approaches tested. From the analysis, it emerges that individual models without decomposition deliver similar performance, with RMSE values around 70 and  $R^2$  near 0.954, suggesting that these models can not accurately capture the characteristics of the time series. Adopting Bi-LSTM significantly improves, with an RMSE of 38.62 and an  $R^2$  of 0.9865, demonstrating a stronger ability to model data variability. The integration of the LSR method further improves performance, reducing RMSE to 37.24 and increasing  $R^2$  to 0.9869. The application of the ICEEMDAN decomposition shows moderate improvements in the ELM and SVM models, with a slight reduction in RMSE. The inclusion of the same technique in Bi-LSTM and LSR models allows for greater predictive accuracy, with RMSE values of 40.69 and 39.75 and a coefficient of determination reaching 0.9844 in the case of the LSR-based model.

The use of VMD decomposition yields a more significant improvement than ICEEMDAN and the models without any decomposition. Applying this technique to the SVM and LSR models produces much more accurate results, with the former achieving an RMSE of 14.29 and an MAE of 0.0181 and the latter recording an RMSE of 23.21, an MAE of 0.0137, and a coefficient of determination of 0.9990. This indicates that combining VMD with LSR represents the most effective approach among those tested, ensuring extremely high predictive accuracy. The differentiated weighting of the most relevant factors through least-squares regression thus significantly improves forecast performance, making this model the best-performing among those analyzed (2.13% error and 99.97%  $R^2$ ).

Regarding the Brasilia site, the error evaluation results for the four models applied to the three datasets are presented in Table 12. The results show that models without decomposition, such as ELM and SVM, perform poorly, with RMSE values of 127.97 and 122.90, respectively, and relatively low  $R^2$  values around 0.88. Using Bi-LSTM significantly improves the forecast, reducing the RMSE to 63.12 and increasing  $R^2$  to 0.97. The LSR-based model provides a further enhancement, achieving an RMSE of 60.40 and an  $R^2$  of 0.9745, suggesting that this technique can optimize the forecasting process.

Applying ICEEMDAN decomposition to the ELM and SVM models does not significantly improve, with the RMSE values remaining high. However, when the same technique is applied to the Bi-LSTM and LSR models, forecasting accuracy improves considerably, with RMSE values dropping to 56.65 and 52.54, respectively, and  $R^2$  increasing to 0.9790 in the case of LSR. These results indicate that decomposition can offer a competitive advantage, but the benefits depend on the model it is combined with.

**Table 11.** Real-time forecasting results at the Tamanrasset site using different methods: (i) without any signal decomposition, (ii) with decomposition via ICEEMDAN, and (iii) with decomposition via VMD. The comparison illustrates the impact of each approach on the prediction accuracy of solar radiation and wind speed.

Model	RMSE	NRMSE	MAE	$R^2$
ELM	69.7311	0.0059	0.0773	0.9540
SVM	70.1124	0.0845	0.0831	0.9535
BI-LSTM	38.6235	0.0004	0.0454	0.9865
<b>LSR</b>	<b>37.2436</b>	<b>0.00035</b>	<b>0.0444</b>	<b>0.9869</b>
ELM-ICEEMDAN	69.6953	0.00083	0.0819	0.9522
SVM-ICEEMDAN	62.0825	0.00055	0.0666	0.9624
Bi-LSTM-ICEEMDAN	40.6926	0.00045	0.0499	0.9838
<b>LSR-ICEEMDAN</b>	<b>39.7556</b>	<b>0.00039</b>	<b>0.0476</b>	<b>0.9844</b>
ELM-VMD	27.3725	0.00031	0.0381	0.9522
SVM-VMD	14.2896	0.00012	0.0181	0.9624
BI-LSTM-VMD	47.0487	0.00013	0.0753	0.9816
<b>LSR-VMD</b>	<b>23.2119</b>	<b>0.00001</b>	<b>0.0137</b>	<b>0.9990</b>

**Table 12.** Real-time forecasting results for the Brasilia site using different models: comparison between predictions without decomposition and those obtained after applying ICEEMDAN and VMD signal decomposition techniques.

Model	RMSE	NRMSE	MAE	$R^2$
ELM	127.97	0.0013	0.1403	0.8802
SVM	122.90	0.0011	0.1200	0.8889
LSTM	63.12	0.0006	0.0812	0.9707
<b>LSR</b>	<b>60.4065</b>	<b>0.00041</b>	<b>0.0789</b>	<b>0.9745</b>
ELM-ICEEMDAN	126.8414	0.00013	0.1451	0.8844
SVM-ICEEMDAN	109.1721	0.0010	0.1042	0.9058
Bi-LSTM-ICEEMDAN	56.6511	0.0006	0.0796	0.9746
<b>LSR-ICEEMDAN</b>	<b>52.5483</b>	<b>0.00041</b>	<b>0.0700</b>	<b>0.9790</b>
ELM-VMD	58.8555	0.00058	0.0674	0.9745
SVM-VMD	26.9754	0.00024	0.0279	0.9943
BI-LSTM-VMD	14.6324	0.00017	0.0213	0.9979
<b>LSR-VMD</b>	<b>9.9046</b>	<b>0.00019</b>	<b>0.0319</b>	<b>0.9956</b>

The integration of VMD decomposition leads to a significant performance improvement compared to ICEEMDAN and models without decomposition. The results show that LSR-VMD is the best-performing model, with an RMSE of 9.90 and an  $R^2$  of 0.9956, confirming this combination's ability to enhance forecast quality drastically. This is followed in order of accuracy by VMD-Bi-LSTM, VMD-SVM, and VMD-ELM, with the first achieving an RMSE of 14.63 and an  $R^2$  of 0.9979. These findings suggest that VMD decomposition enables a more effective separation of the signal's informative components, significantly improving prediction accuracy, especially when combined with advanced methods such as Bi-LSTM and LSR.

An analysis of the results obtained for the Colorado site (see Table 13) reveals that the use of VMD led to a significant improvement in performance compared to models that do not include a decomposition phase. In particular, the SVM-VMD and LSR-VMD models outperformed all others, with RMSE values of 2.2942 and 2.1823, respectively, and a coefficient of determination of 0.9997 in both cases. This outcome highlights how combining VMD with advanced regression strategies can capture time series variations more effectively, reduce error, and enhance predictive capability.

**Table 13.** Real-time forecasting results for the Colorado site using different predictive methods. The comparison includes models without signal decomposition and those enhanced through ICEEMDAN- and VMD-based decomposition techniques.

Model	RMSE	NRMSE	MAE	$R^2$
ELM	20.284	0.0014	0.0753	0.9801
SVM	17.701	0.0009	0.0609	0.9848
Bi-LSTM	12.8850	0.0009	0.0507	0.9920
<b>LSR</b>	<b>11.4794</b>	<b>0.0005</b>	<b>0.0423</b>	<b>0.9937</b>
ELM-ICEEMDAN	14.4662	0.00089	0.0465	0.9885
SVM-ICEEMDAN	20.7725	0.00540	0.0754	0.9764
LSTM-ICEEMDAN	11.5208	0.00086	0.0414	0.9905
<b>LSR-ICEEMDAN</b>	<b>10.6804</b>	<b>0.00058</b>	<b>0.0400</b>	<b>0.9941</b>
ELM-VMD	10.2433	0.00077	0.0419	0.9943
SVM-VMD	2.2942	0.00014	0.0088	0.9997
Bi-LSTM-VMD	4.6764	0.00035	0.0148	0.9888
<b>LSR-VMD</b>	<b>2.1823</b>	<b>0.00012</b>	<b>0.0083</b>	<b>0.9997</b>

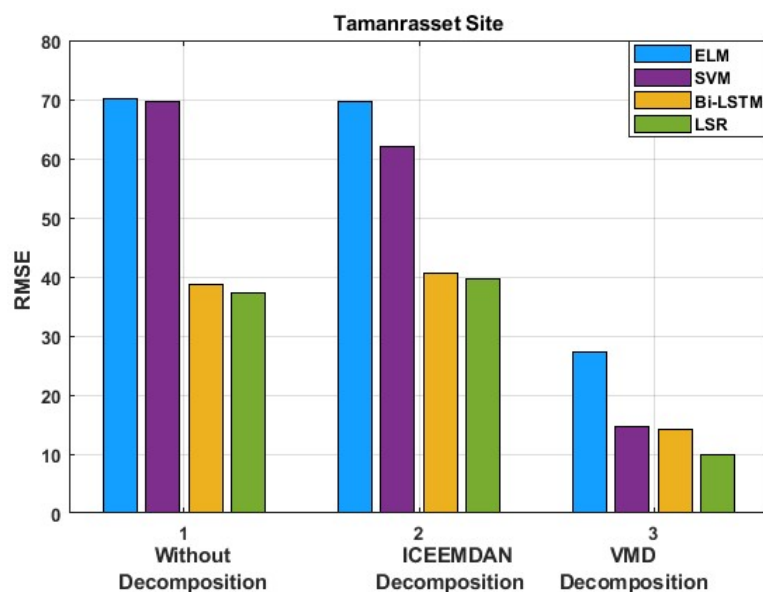
The ICEEMDAN-based approach showed improvements over models without decomposition but did not reach the performance level of VMD. The LSR-ICEEMDAN model achieved an RMSE of 10.6804 and an  $R^2$  of 0.9941, confirming good predictive ability, though still lower than the VMD-based models. Similarly, the Bi-LSTM-ICEEMDAN model showed decent accuracy, with an RMSE of 11.5208 and an MAE of 0.0414, although inferior to the VMD alternatives.

Among the models without decomposition, Bi-LSTM performed better than ELM and SVM, with an RMSE of 12.8850 and an  $R^2$  of 0.9920. However, the integration of VMD led to a substantial improvement in accuracy, significantly reducing prediction error. This demonstrates that time series decomposition is crucial in optimizing forecast performance.

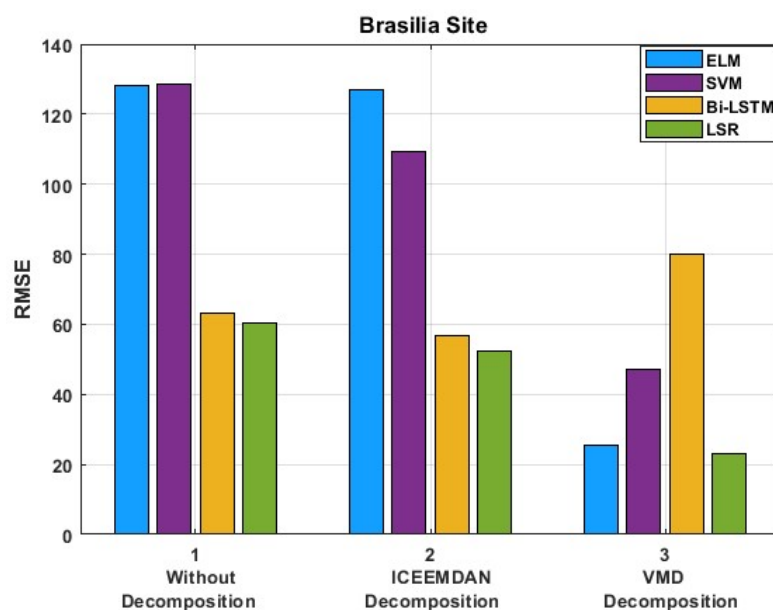
Overall, the results indicate that the most effective method for the Colorado site is the combination of VMD with LSR, which achieved the lowest error rate and the highest  $R^2$  value. The effectiveness of VMD decomposition is attributed to its ability to isolate the most informative modal components, eliminate noise, and facilitate the extraction of relevant patterns for prediction. Although ICEEMDAN proved helpful in improving the separation of frequency components compared to baseline models, its contribution was less substantial than that of VMD.

These findings confirm the importance of integrating advanced decomposition methods into predictive models to enhance forecast quality and improve uncertainty handling in energy time series.

Figures 20–22 show the RMSE for each model at each site, clearly highlighting that the SVM-ELM-Bi-LSTM-LSR model outperforms the best individual model (Bi-LSTM). This is due to the aggregation strategy, which sums the products of the weights and the forecasts generated by the three models, assigning greater weight to the higher-performing ones. Additionally, Bi-LSTM enhances the model's ability to retain long-term information, while the attention mechanism highlights the most critical features through the dynamic weighting of details after the Bi-LSTM output. In summary, the prediction accuracy of the combined model surpasses that of the other models analyzed, achieving better results than individual approaches.

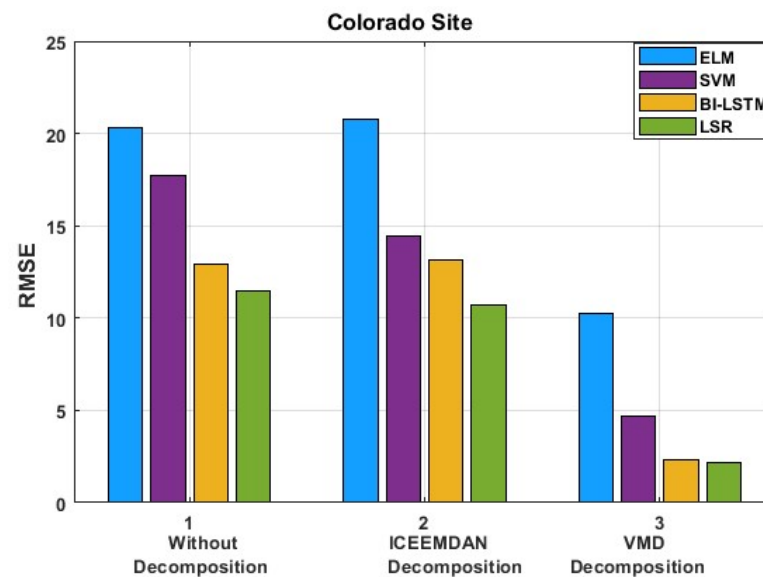


**Figure 20.** Comparison of RMSE values obtained for the Tamanrasset site before and after applying VMD and ICEEMDAN decomposition techniques. The figure highlights the improvement in forecasting accuracy achieved through the use of signal decomposition prior to model training.



**Figure 21.** Comparison of RMSE values obtained for the Brasilia site before and after applying the VMD and ICEEMDAN decomposition methods. The figure illustrates the impact of each preprocessing technique on the forecasting accuracy of the proposed model.

**Remark 12.** To evaluate the effectiveness of the proposed forecasting framework, a systematic comparison was conducted with several baseline models commonly used in the literature. Specifically, the following approaches were implemented and tested: (i) the ARIMA model, representative of traditional statistical time series methods; (ii) a Bi-LSTM model applied directly to the raw data, without any decomposition or entropy-based selection; and (iii) a standard SVM model applied without any pre-processing phase. The results clearly demonstrate that the proposed hybrid framework, combining decomposition (VMD or ICEEMDAN), SE-based selection, and LSR, significantly outperforms all baseline models. In particular, the proposed method achieves up to a 30% reduction in forecasting error, with a coefficient of determination  $R^2$  approaching 0.999.



**Figure 22.** Root mean square error (RMSE) values for the Colorado dataset, comparing forecasting performance before and after the application of VMD and ICEEMDAN signal decomposition techniques. The results highlight the improvement in prediction accuracy achieved through the proposed pre-processing methods.

## 8. Conclusions and Perspectives

This study proposed an innovative hybrid framework for the forecasting of renewable energy sources, focusing on global solar radiation and wind speed. The methodology is based on a one-stage signal decomposition strategy that integrates variational mode decomposition (VMD) and improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), followed by an entropy-based selection process using sample entropy (SE). This approach allows for the extraction and retention of the most informative components of the original signals, effectively reducing noise and improving data quality before prediction.

In the predictive stage, three different machine learning models—Bi-LSTM, extreme learning machine (ELM), and support vector regression (SVR)—were employed to capture the nonlinear dynamics and temporal dependencies present in the decomposed data. These individual predictions were then optimally combined using a least-squares regression (LSR) strategy, enhancing the overall forecasting accuracy and robustness of the system. The framework was applied to real-world datasets from three distinct locations—Tamanrasset (Algeria), Brasilia (Brazil), and Colorado (USA)—which present heterogeneous climatic conditions, providing a comprehensive validation of the model's generalization capability.

The experimental results showed that the proposed model achieves a root mean square error (RMSE) as low as 2.18 and a coefficient of determination ( $R^2$ ) approaching 0.999, confirming the model's high precision. Compared to traditional single-model approaches, the framework demonstrated up to a 30% reduction in forecasting error. In terms of computational efficiency, the average training time for the Bi-LSTM component ranged between 98 and 158 s depending on the dataset, while the LSR combination phase required less than 0.5 s, highlighting the suitability of the approach for operational scenarios.

**Author Contributions:** Conceptualization, M.V. and F.L.F.; methodology, M.V., F.L.F. and G.A.; software, N.Z., H.M. and Z.Z.; validation, M.V., G.A. and Z.Z.; formal analysis, N.Z., H.M. and F.L.F.; investigation, M.V. and F.L.F.; resources, F.L.F. and M.V.; data curation N.Z., H.M. and Z.Z.; writing—original draft preparation, M.V. and G.A.; writing—review and editing, M.V. and G.A.;

supervision, M.V., F.L.F. and G.A.; project administration, M.V.; funding acquisition, M.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the Next Generation EU—Italian NRRP, Mission 4, Component 2, Investment 1.5, call for the creation and strengthening of ‘Innovation Ecosystems’, building ‘Territorial R&D Leaders’ (Directorial Decree n. 2021/3277)—project Tech4You—Technologies for climate change adaptation and quality of life improvement, n. ECS0000009 (particularly, action 9 of Spoke 2—Goal 2.1—Pilot Project 1). This work reflects only the authors’ views and opinions; neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

**Data Availability Statement:** Data are contained within the article.

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

## Abbreviations

The following abbreviations are used in this manuscript:

VMD	Variational Mode Decomposition
ICEEMDAN	Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
NSD	Numerical Weather Prediction
EMD	Empirical Mode Decomposition
EEMD	Ensemble Empirical Mode Decomposition
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
ENN	Elman Neural Network
ELM	Extreme Machine Learning
SARIMA	Seasonal Autoregressive Integrated Moving Average model
RES	Renewable Energy Sources
GHI	Global Horizontal Irradiance
WS	Wind Speed
Bi-LSTM	Long Short-Term Memory
LSR	Least-Squares Regression

## References

- Ioannou, A.; Angus, A.; Brennan, F. Risk-Based Methods for Sustainable Energy System Planning: A Review. *Renew. Sustain. Energy Rev.* **2017**, *74*, 602–615. [\[CrossRef\]](#)
- León Gómez, J.C.; De León Aldaco, S.E.; Aguayo Alquicira, J. A Review of Hybrid Renewable Energy Systems: Architectures, Battery Systems, and Optimization Techniques. *Eng* **2023**, *4*, 1446–1467. [\[CrossRef\]](#)
- Rouhandeh, M.; Ahmadi, A.; Mirhosseini, M.; Alirezaei, R. Economic energy Supply Using Renewable Sources such as Solar and Wind in Hard-to-reach Areas of Iran with two Different Geographical Locations. *Energy Strategy Rev.* **2024**, *55*, 101494. [\[CrossRef\]](#)
- Suliman, F.E.M. Solar- and Wind-Energy Utilization in the Kingdom of Saudi Arabia: A Comprehensive Review. *Energies* **2024**, *17*, 1894. [\[CrossRef\]](#)
- Barbhuiya, S.; Kanavaris, F.; Das, B.B.; Idrees, M. Decarbonising Cement and Concrete Production: Strategies, Challenges and Pathways for Sustainable Development. *J. Build. Eng.* **2024**, *86*, 108861. [\[CrossRef\]](#)
- Wu, Y.; Wang, J. A Novel Hybrid Model Based on Artificial Neural Networks for Solar Radiation Prediction. *Renew. Energy* **2016**, *89*, 268–284. [\[CrossRef\]](#)
- Ur Rehman, Z.; Tariq, S.; ul Haq, Z.; Khan, M. Impact of Meteorological Parameters on Aerosol Optical Depth and Particulate Matter in Lahore. *Acta Geophys.* **2024**, *72*, 1377–1395. [\[CrossRef\]](#)
- Islam, M.M.; Yu, T.; Giannoccaro, G.; Mi, Y.; La Scala, M.; Rajabi, M.N.; Wang, J. Improving Reliability and Stability of the Power Systems: A Comprehensive Review on the Role of Energy Storage Systems to Enhance Flexibility. *IEEE Access* **2024**, *12*, 152738–152765. [\[CrossRef\]](#)
- Wu, J.; He, D.; Li, J.; Miao, J.; Li, X.; Li, H.; Shan, S. Temporal Multi-Resolution Hypergraph Attention Network for Remaining Useful Life Prediction of Rolling Bearings. *Reliab. Eng. Syst. Saf.* **2024**, *247*, 110143. [\[CrossRef\]](#)
- Bassey, K.E.; Juliet, A.R.; Stephen, A.O. AI-Enhanced Lifecycle Assessment of Renewable Energy Systems. *Eng. Sci. Technol. J.* **2024**, *5*, 2082–2099. [\[CrossRef\]](#)
- Hamdan, A.; Ibekwe, K.I.; Ilojiyana, V.I.; Sonko, S.; Etukudoh, E.A. AI in Renewable Energy: A Review of Predictive Maintenance and Energy Optimization. *Int. J. Sci. Res. Arch.* **2024**, *11*, 718–729. [\[CrossRef\]](#)

12. Zhang, H.; Zhang, M.; Yi, R.; Liu, Y.; Wen, Q.H.; Meng, X. Growing Importance of Micro-Meteorology in the New Power System: Review, Analysis and Case Study. *Energies* **2024**, *17*, 1365. [[CrossRef](#)]
13. Chisale, S.W.; Lee, H.S. Comprehensive Onshore Wind Energy Assessment in Malawi Based on the WRF Downscaling with ERA5 Reanalysis Data, Optimal Site Selection, and Energy Production. *Energy Convers. Manag. X* **2024**, *22*, 100608. [[CrossRef](#)]
14. Koeva, D.; Kutkarska, R.; Zinoviev, V. High Penetration of Renewable Energy Sources and Power Market Formation for Countries in Energy Transition: Assessment via Price Analysis and Energy Forecasting. *Energies* **2023**, *16*, 7788. [[CrossRef](#)]
15. Ascione, G.; Bufalo, M.; Orlando, G.; Quadrini, R. Balancing the Grid: Mitigating the Effects of Renewable Energy in Italy via Skew Modeling and Forecasting. *Ann. Oper. Res.* **2024**, 1–39. [[CrossRef](#)]
16. Unsal, D.B.; Aksoz, A.; Oyucu, S.; Guerrero, J.M.; Guler, M. A Comparative Study of AI Methods on Renewable Energy Prediction for Smart Grids: Case of Turkey. *Sustainability* **2024**, *16*, 2894. [[CrossRef](#)]
17. Al Hadi, F.M.; Aly, H.H. Harmonics Forecasting of Renewable Energy System using Hybrid Model based on LSTM and ANFIS. *IEEE Access* **2024**, *12*, 50966–50985. [[CrossRef](#)]
18. Lay-Ekuakille, A.; Palamara, I.; Caratelli, D.; Morabito, F.C. Experimental Infrared Measurements for Hydrocarbon Pollutant Determination in Subterranean Waters. *Rev. Sci. Instrum.* **2013**, *84*, 015111. [[CrossRef](#)]
19. Yang, H.; Wu, Q.; Li, G. A Multi-Stage Forecasting System for Daily Ocean Tidal Energy Based on Secondary Decomposition, Optimized Gate Recurrent Unit and Error Correction. *J. Clean. Prod.* **2024**, *449*, 141303. [[CrossRef](#)]
20. Ding, S.; Zhang, H.; Tao, Z.; Li, R. Integrating data decomposition and machine learning methods: An empirical proposition and analysis for renewable energy generation forecasting. *Expert Syst. Appl.* **2022**, *204*, 117635. [[CrossRef](#)]
21. Pu, S.; Li, Z.; Wan, H.; Chen, Y. A Hybrid Prediction Model for Photovoltaic Power Generation Based on Information Entropy. *IET Gener. Transm. Distrib.* **2021**, *15*, 436–455. [[CrossRef](#)]
22. Niu, Q.; You, M.; Yang, Z.; Zhang, Y. Economic Emission Dispatch Considering Renewable Energy Resources—A Multi-Objective Cross Entropy Optimization Approach. *Sustainability* **2021**, *13*, 5386. [[CrossRef](#)]
23. Montesinos, L.; Castaldo, R.; Pecchia, L. On the Use of Approximate Entropy and Sample Entropy with Centre of Pressure Time-Series. *J. Neuroeng. Rehabil.* **2018**, *15*, 116. [[CrossRef](#)] [[PubMed](#)]
24. Pande, C.B.; Kushwaha, N.L.; Alawi, O.A.; Sammen, S.S.; Sidek, L.M.; Yaseen, Z.M.; Pal, S.C.; Katipoğlu, O.M. Daily Scale Air Quality Index Forecasting Using Bidirectional Recurrent Neural Networks: Case Study of Delhi, India. *Environ. Pollut.* **2024**, *351*, 124040. [[CrossRef](#)]
25. Barjasteh, A.; Ghafouri, S.H.; Hashemi, M. A Hybrid Model Based on Discrete Wavelet Transform (DWT) and Bidirectional Recurrent Neural Networks for Wind Speed Prediction. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107340. [[CrossRef](#)]
26. Samadi-Koucheksaraee, A.; Chu, X. Development of a Novel Modeling Framework Based on Weighted Kernel Extreme Learning Machine and Ridge Regression for Streamflow Forecasting. *Sci. Rep.* **2024**, *14*, 30910. [[CrossRef](#)]
27. Bonakdari, H.; Ebtehaj, I.; Samui, P.; Gharabaghi, B. Lake Water-Level Fluctuations Forecasting Using Minimax Probability Machine Regression, Relevance Vector Machine, Gaussian Process Regression, and Extreme Learning Machine. *Water Resour. Manag.* **2019**, *33*, 3965–3984. [[CrossRef](#)]
28. Ahmed, A.; Zhang, J.; Wang, B. Hybrid CNN-LSTM Deep Learning Model for Direct Solar Irradiance Forecasting. *Sci. Rep.* **2025**, *15*, 5891. [[CrossRef](#)]
29. Wang, Y.; Zhang, Y.; Li, Z.; Huang, S. Adaptive Wind Speed Calibration Using C-LSTM for Enhanced Wind Power Forecasting. *Sci. Rep.* **2025**, *15*, 3489. [[CrossRef](#)]
30. Sadeghian, A.; Alimardani, M.; Zare, M. AROA-LSTM: Wind Power Forecasting Using Attraction-Repulsion Optimization in Deep Learning Frameworks. *Energy Rep.* **2025**, *10*, 312–324. [[CrossRef](#)]
31. Xu, Z.; Zhao, H.; Xu, C.; Shi, H.; Xu, J.; Wang, Z. A Novel Wind Power Prediction Model That Considers Multi-Scale Variable Relationships and Temporal Dependencies. *Electronics* **2024**, *13*, 3710. [[CrossRef](#)]
32. Malakar, S.; Goswami, S.; Chakrabarti, A.; Ganguli, B. A Novel Denoising Technique and Deep Learning Based Hybrid Wind Speed Forecasting Model for Variable Terrain Conditions. *arXiv* **2024**, arXiv:2408.15554.
33. Zemouri, N.; Bouzgou, H.; Gueymard, C.A. Multimodel Ensemble Approach for Hourly Global Solar Irradiation Forecasting. *Eur. Phys. J. Plus* **2019**, *134*, 12. [[CrossRef](#)]
34. Ibrahim, A.; Li, M.; Yu, T.; Tian, H.; Wang, H.; Liu, Y. Wind Speed Ensemble Forecasting Based on Deep Learning Using Adaptive Dynamic Optimization Algorithm. *IEEE Access* **2021**, *9*, 125787–125804. [[CrossRef](#)]
35. Makubiyane, K.; Maposa, D. Forecasting Short-and Long-Term Wind Speed in Limpopo Province Using Machine Learning and Extreme Value Theory. *Forecasting* **2024**, *6*, 885–907. [[CrossRef](#)]
36. Xu, C.; Chen, Y.; Zhao, X.; Song, W.; Li, X. Prediction of seawater pH by bidirectional gated recurrent neural network with attention under phase space reconstruction: Case study of the coastal waters of Beihai, China. *Acta Oceanol. Sin.* **2023**, *42*, 97–107. [[CrossRef](#)]
37. Almarzooqi, A.M.; Maalouf, M.; El-Fouly, T.H.M.; Katzourakis, V.E.; El Moursi, M.S.; Chrysikopoulos, C.V. A Hybrid Machine-Learning Model for Solar Irradiance Forecasting. *Clean Energy* **2024**, *8*, 100–110. [[CrossRef](#)]

38. Poggi, P.; Muselli, M.; Notton, G.; Cristofari, C.; Louche, A. Forecasting and Simulating Wind Speed in Corsica by Using an Autoregressive Model. *Energy Convers. Manag.* **2003**, *44*, 3177–3196. [[CrossRef](#)]
39. Demolli, H.; Dokuz, A.S.; Ecemis, A.; Gokcek, M. Wind Power Forecasting Based on Daily Wind Speed Data Using Machine Learning Algorithms. *Energy Convers. Manag.* **2019**, *198*, 111823. [[CrossRef](#)]
40. Sun, L.; Lan, Y.; Sun, X.; Liang, X.; Wang, J.; Su, Y.; He, Y.; Xia, D. Deterministic Forecasting and Probabilistic Post-Processing of Short-Term Wind Speed Using Statistical Methods. *J. Geophys. Res. Atmos.* **2024**, *129*, e2023JD040134. [[CrossRef](#)]
41. Zhao, X.; Liu, J.; Yu, D.; Chang, J. One-Day-Ahead Probabilistic Wind Speed Forecast Based on Optimized Numerical Weather Prediction Data. *Energy Convers. Manag.* **2018**, *164*, 560–569. [[CrossRef](#)]
42. Mishra, S.P.; Dash, P.K. Short Term Wind Speed Prediction Using Multiple Kernel Pseudo Inverse Neural Network. *Int. J. Autom. Comput.* **2018**, *15*, 66–83. [[CrossRef](#)]
43. Wang, Z.; Wang, F.; Su, S. Solar Irradiance Short-Term Prediction Model Based on BP Neural Network. *Energy Procedia* **2011**, *12*, 488–494. [[CrossRef](#)]
44. Barros, L.A.; Tanta, M.; Martins, A.P.; Pinto, J.G. *Sustainable Energy for Smart Cities*; Springer: Berlin/Heidelberg, Germany, 2021; Volume 375. [[CrossRef](#)]
45. Arati, D.C.; Menon, P.S.; Velayudhan, J.; Poornachandran, P.; Raj, A.K.; OK, S. Enhancing Wind Power Prediction through Machine Learning. In Proceedings of the 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP), IEEE, Andhra Pradesh, India, 22–24 February 2024; pp. 1–5.
46. Aksu, G.; Güzeller, C.O.; Eser, M.T. The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model. *Int. J. Assess. Tools Educ.* **2019**, *6*, 170–192. [[CrossRef](#)]
47. Versaci, M.; La Foresta, F. Fuzzy Approach for Managing Renewable Energy Flows for DC-Microgrid with Composite PV-WT Generators and Energy Storage System. *Energies* **2024**, *17*, 402. [[CrossRef](#)]
48. Versaci, M.; Angiulli, G.; La Foresta, F.; Crucitti, P.; Laganá, F.; Pellicanó, D.; Palumbo, A. Innovative Soft Computing Techniques for the Evaluation of the Mechanical Stress State of Steel Plates. In *International Conference on Applied Intelligence and Informatics*; Springer Nature: Cham, Switzerland, 2021; pp. 14–28.
49. Versaci, M.; La Foresta, F.; Laganà, F.; Morabito, F.C. Stand-Alone DC-MSs & TS Fuzzy Systems for Regenerative Urban Design. In *International Symposium: New Metropolitan Perspectives*; Springer: Cham, Switzerland, 2024; pp. 36–48.
50. Ouacel, A.; M'hamdi, B.; Benaissa, T.; Amari, A.; Pietrafesa, M.; Versaci, M. Maximum Power Point Tracking Based on Improved Sliding Mode Control. *Eur. Phys. J. Plus* **2025**, *140*, 340. [[CrossRef](#)]
51. Postorino, M.N.; Versaci, M. A neuro-fuzzy approach to simulate the user mode choice behaviour in a travel decision framework. *Int. J. Model. Simul.* **2008**, *28*, 64–71. [[CrossRef](#)]
52. Abedinia, O.; Amjady, N. Short-Term Wind Power Prediction Based on Hybrid Neural Network and Chaotic Shark Smell Optimization. *Int. J. Precis. Eng. Manuf. Green Technol.* **2015**, *2*, 245–254. [[CrossRef](#)]
53. Xiao, L.; Qian, F.; Shao, W. Multi-Step Wind Speed Forecasting Based on a Hybrid Forecasting Architecture and an Improved Bat Algorithm. *Energy Convers. Manag.* **2017**, *143*, 410–430. [[CrossRef](#)]
54. Bae, K.Y.; Jang, H.S.; Sung, D.K. Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis. *IEEE Trans. Power Syst.* **2017**, *32*, 935–945. [[CrossRef](#)]
55. Yu, C.; Li, Y.; Zhang, M. Comparative Study on Three New Hybrid Models Using Elman Neural Network and Empirical Mode Decomposition Based Technologies Improved by Singular Spectrum Analysis for Hour-Ahead Wind Speed Forecasting. *Energy Convers. Manag.* **2017**, *147*, 75–85. [[CrossRef](#)]
56. Peng, T.; Zhou, J.; Zhang, C.; Zheng, Y. Multi-Step Ahead Wind Speed Forecasting Using a Hybrid Model Based on Two-Stage Decomposition Technique and AdaBoost-Extreme Learning Machine. *Energy Convers. Manag.* **2017**, *153*, 589–602. [[CrossRef](#)]
57. Haddad, M.; Nicod, J.; Mainassara, Y.B.; Rabehasaina, L.; Al Masry, Z.; Péra, M. Wind and Solar Forecasting for Renewable Energy System Using SARIMA-Based Model. In Proceedings of the 2019 IEEE PES Conference on Innovative Smart Grid Technologies Latin America (ISGT LATAM), Gramado, Brazil, 15–18 September 2019; pp. 1–6.
58. Gupta, A.; Gupta, K.; Saroha, S. A Review and Evaluation of Solar Forecasting Technologies. *Mater. Today Proc.* **2021**, *47*, 2420–2425. [[CrossRef](#)]
59. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [[CrossRef](#)]
60. Wang, Y.; Liu, F.; Jiang, Z.; He, S.; Mo, Q. Complex Variational Mode Decomposition for Signal Processing Applications. *Mech. Syst. Signal Process.* **2017**, *86*, 75–85. [[CrossRef](#)]
61. Zayed, A.I. Hilbert Transform Associated with the Fractional Fourier Transform. *IEEE Signal Process. Lett.* **1998**, *5*, 206–208. [[CrossRef](#)]
62. Wang, Y.; Markert, R. Filter bank property of variational mode decomposition and its applications. *Signal Process.* **2016**, *120*, 509–521. [[CrossRef](#)]
63. Birgin, Ernesto G., Martínez, J.M. *Practical Augmented Lagrangian Methods for Constrained Optimization*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2014.

64. Li, W.; Shi, Q.; Sibtain, M.; Li, D.; Mbanze, D.E. A Hybrid Forecasting Model for Short-Term Power Load Based on Sample Entropy, Two-Phase Decomposition and Whale Algorithm Optimized Support Vector Regression. *IEEE Access* **2020**, *8*, 166907–166921. [[CrossRef](#)]
65. Yang, J.; Guan, H.; Ma, X.; Zhang, Y.; Lu, Y. Rapid Detection of Corn Moisture Content Based on Improved ICEEMDAN Algorithm Combined with TCN-BiGRU Model. *Food Chem.* **2025**, *465*, 142133. [[CrossRef](#)]
66. Lu, B.; Li, Z.; Zhang, X. ICEEMDAN–RPE–AITD Algorithm for Magnetic Field Signals of Magnetic Targets. *Sci. Rep.* **2025**, *15*, 6509. [[CrossRef](#)]
67. Sun, W.; Wang, Y. Short-Term Wind Speed Forecasting Based on Fast Ensemble Empirical Mode Decomposition, Phase Space Reconstruction, Sample Entropy and Improved Back-Propagation Neural Network. *Energy Convers. Manag.* **2018**, *157*, 1–12. [[CrossRef](#)]
68. Xia, X.; Wang, X. A Novel Hybrid Model for Short-Term Wind Speed Forecasting Based on Twice Decomposition, PSR, and IMVO-ELM. *Complexity* **2022**, *2022*, 4014048. [[CrossRef](#)]
69. Richman, J.S.; Moorman, J.R. Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]
70. Liu, H.; Wang, N.; Fang, H.; Yu, X.; Du, W. Identification of the Number of Leaks in Water Supply Pipes Based on Wavelet Scattering Network and Bi-LSTM Model with Bayesian Optimization. *Measurement* **2025**, *243*, 116348. [[CrossRef](#)]
71. Tasnim, J.; Rahman, M.A.; Rafi, M.S.A.; Talukder, M.A.; Hasan, M.K. Generalized Real-Time State of Health Estimation for Lithium-Ion Batteries Using Simulation-Augmented Multi-Objective Dual-Stream Fusion of Multi-Bi-LSTM-Attention. *e-Prime-Adv. Electr. Eng. Electron. Energy* **2025**, *11*, 100870. [[CrossRef](#)]
72. Pramanik, A.; Sarker, S.; Sarkar, S.; Pal, S.K. Real-Time Fall Detection on Road Using Transfer Learning-Based Granulated Bi-LSTM. *Knowl.-Based Syst.* **2025**, *311*, 113038. [[CrossRef](#)]
73. Kim, W.S.; Kim, T.M.; Cho, S.G.; Jarque, I.; Iskierka-Jażdżewska, E.; Poon, L.M.; Walewski, J. Odronextamab Monotherapy in Patients with Relapsed/Refractory Diffuse Large B Cell Lymphoma: Primary Efficacy and Safety Analysis in Phase 2 ELM-2 Trial. *Nat. Cancer* **2025**, *6*, 528–539. [[CrossRef](#)]
74. Kulagin, V.; Ding, H.; Huang, Q.; Razmjoo, N. An Improved Version of Firebug Swarm Optimization Algorithm for Optimizing Alex/ELM Network Kidney Stone Detection. *Biomed. Signal Process. Control* **2025**, *99*, 106898.
75. Gasparyan, Y. Numerical Simulation of Deuterium Retention in Tungsten under ELM-Like Conditions. *J. Nucl. Mater.* **2025**, *603*, 155370.
76. Drucker, H.; Surges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support Vector Regression Machines. *Adv. Neural Inf. Process. Syst.* **1997**, *9*, 155–161.
77. Li, L.L.; Cen, Z.Y.; Tseng, M.L.; Shen, Q.; Ali, M.H. Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic—Support vector regression machine. *J. Clean. Prod.* **2021**, *279*, 123739. [[CrossRef](#)]
78. Hu, Q.; Zhang, S.; Xie, Z.; Mi, J.; Wan, J. Noise Model Based  $\nu$ -Support Vector Regression with Its Application to Short-Term Wind Speed Forecasting. *Neural Netw.* **2014**, *57*, 1–11. [[CrossRef](#)] [[PubMed](#)]
79. Chen, K.; Yu, J. Short-Term Wind Speed Prediction Using an Unscented Kalman Filter Based State-Space Support Vector Regression Approach. *Appl. Energy* **2014**, *113*, 690–705. [[CrossRef](#)]
80. Sturges, J.; Yang, L.; Shafaei, S.; Lan, C. Efficient Data-Dependent Random Projection for Least Square Regressions. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Hyderabad, India, 6–11 April 2025.
81. Yi, M.; Chen, W.; Liu, J. Total Partial Least Square Regression and Its Application in Infrared Spectra Quantitative Analysis. *Measurement* **2025**, *226*, 116794. [[CrossRef](#)]
82. Sai Kalyan, B.H.; Nachiar, S.S. Predicting stakeholder perspective of alternative dispute resolution in the construction industry using ordinary least square regression. *Organ. Technol. Manag. Constr.* **2025**, *17*, 1. [[CrossRef](#)]

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