Smart Urban Wind Power Forecasting: Integrating Weibull Distribution, Recurrent Neural Networks, and Numerical Weather Prediction

Navid Shirzadi 1, Fuzhan Nasiri 1,*, Ramanunni Parakkal Menon 1, Pilar Monsalvete 1, Anton Kaifel 2 and Ursula Eicker 1

1 Gina Cody School of Engineering and Computer Science, Concordia University, 1455 Boulevard de Maisonneuve, Montreal, QC H3G 1M8, Canada; navid.shirzadi@concordia.ca (N.S.); ramanunni.parakkal.menon@concordia.ca (R.P.M.); pilar.monsalvete@concordia.ca (P.M.); ursula.eicker@concordia.ca (U.E.)
2 Centre for Solar Energy and Hydrogen Research (ZSW), Meitnerstr. 1, 70563 Stuttgart, Germany; anton.kaifel@zsw-bw.de
* Correspondence: fuzhan.nasiri@concordia.ca

Abstract: The design, operational planning, and integration of wind power plants with other renewables and the grid face challenges attributed to the intermittent nature of wind power generation. Addressing this issue necessitates the development of a smart wind power (and in particular wind speed) forecasting approach. This is a complex task due to substantial fluctuations in wind speed. To overcome the inherent stochastic nature of wind speed and mitigate related challenges, traditionally, numerical weather prediction (NWP) models are employed for wind speed forecasting. However, the applicability of NWP models is limited to short-term forecasting due to their computational constraints. In this study, a hybrid AI-based approach is proposed to improve forecast accuracy over a 48 h horizon for the city of Montreal. The results demonstrate that by integrating the probability distribution of wind speed with a deep learning model, the forecasted values align closely with the observed values in terms of seasonality and trend, exhibiting enhanced accuracy. Evaluation metrics reveal a substantial reduction in the root mean squared error (13–31%) across three prediction horizons (summer, fall, and winter) compared to a single long, short-term memory model. Furthermore, integrating the improved model with the numerical weather prediction model yields increased accuracy and decreased error compared to the LSTM–Weibull model.

Keywords: wind power generation; wind speed forecasting; deep learning; Weibull distribution; numerical weather prediction; smart cities

1. Introduction

According to the World Wind Energy Association, the global wind power generation achieved a new record of 744 Gigawatts by the end of 2020, with an addition of 93 Gigawatts [1]. However, fluctuations in wind power generation lead to a lack of reliability, posing significant challenges and uncertainties for control systems and operators in ensuring a stable power supply [2]. Hence, the importance of reliable and precise wind forecasting cannot be overstated, as it plays a crucial role in various applications, including load following, unit commitment, scheduling, economic viability, and the design and operational planning of renewable energy systems. Nonetheless, the volatile and intermittent nature of wind speed presents a formidable obstacle in achieving accurate short-term predictions.

Various methods exist for wind prediction, which can be broadly classified into four main groups: (1) physical models, (2) spatial correlation models, (3) conventional statistical
models, and (4) artificial intelligence (AI) models [3]. Physical models utilize meteorological data, including temperature and physical characteristics, and are commonly employed for large-scale weather prediction [4]. On the other hand, spatial correlation models rely on the predicted wind speeds of nearby sites to estimate wind speeds at new locations.

In recent times, both conventional statistical models and AI models have gained significant popularity for intraday wind speed predictions, particularly in the context of the design and operational planning of integrated renewable energy systems and wind farms. These models have been extensively employed to enhance the accuracy of forecasting in such applications.

Wang et al. [4] introduced a hybrid model consisting of an autoregressive moving average (ARMA) model and a bivariate fuzzy time series model to predict daily wind speed in Hainan province, China. The findings of their study reveal that, compared to conventional models such as ARMA and ARIMA, the hybrid model significantly reduces the mean absolute percentage error (MAPE) for day-ahead wind speed forecasting. The MAPE for the conventional models across four different sites ranged from 18.15% to 22.08%. In contrast, the hybrid model achieved an error range of 16.64% to 18.29%, showcasing its improved performance in wind speed prediction.

In 2017, Yatiyana et al. [5] presented a statistical model utilizing autoregressive integrated moving average (ARIMA) to predict wind speed and direction in Western Australia in 2017. The selection of this method was motivated by its shorter response time. Their findings demonstrated a mean absolute percentage error (MAPE) of 4.9% for wind speed prediction at a 6 h lead time and a MAPE of 15.6% for wind direction forecasting with a 7-day lead time. However, the integration of these two models into a single model for enhanced overall accuracy was not reported, and the ARIMA method’s ability to capture wind speed fluctuations was not explained.

In another study, the application of fractional ARIMA for day-ahead and two-day-ahead wind speed forecasting was explored [6]. The results exhibited a significant reduction in error and improvement in accuracy compared to persistence methods. Several studies have employed seasonal autoregressive integrated moving average (SARIMA) models to account for the seasonality of training data. Wang et al. [7] used SARIMA for daily and monthly wind speed forecasting in four sites in Northwestern China. To enhance accuracy, they hybridized SARIMA with an extreme learning machine (ELM) and Ljung–Box Q-test (LBQ), considering the nonlinearity and non-stationarity inherent in wind speed data. In one of the sites, the mean daily forecast results indicated approximately 34% MAPE for the single SARIMA method, whereas their proposed hybrid model achieved an error of about 14%, demonstrating a significant improvement in accuracy.

In addition to the mentioned methods, researchers have explored the use of fuzzy theory [8,9] and machine learning techniques, such as support vector machine (SVM) [10,11], for day-ahead wind forecasting. However, artificial neural network (ANN) has garnered significant attention and is widely employed for wind speed forecasting, either as a standalone model or as part of a hybrid approach in combination with other models, such as statistical models. In a study by [12], a comparison was made between ANN, autoregressive integrated moving average (ARIMA), and a hybrid model combining ARIMA and ANN for wind speed forecasting in three regions of India. The results indicated that the hybrid model exhibited improved performance in wind speed prediction, regardless of the linear or non-linear behavior of the wind speed. Although the hybrid model demonstrated significantly lower error compared to the ANN-only model, the mean absolute percentage error (MAPE) of the hybrid model forecasts (ranging from 18% to 25%) for various lead times (1 h, 3 h, 8 h, and 24 h) still remained relatively high.

In recent years, deep learning techniques, including recurrent neural networks (RNNs), Elman neural networks, and convolutional neural networks (CNNs), have gained significant attention in time series forecasting due to their ability to handle sequential data effectively [13–16]. Liu et al. [17] introduced a hybrid model combining an Elman neural network with a long short-term memory (LSTM) network for wind speed
forecasting. Their findings indicate that LSTM is suitable for predicting non-stationary wind speeds, and their proposed hybrid model achieved reasonable accuracy in forecasting. In another study by Wang et al. [14], wind power forecasting was performed using a CNN model. The results demonstrated the adequacy of the proposed CNN model for wind power prediction.

Although AI and statistical methods generally yield satisfactory results in various forecasting horizons (short term, medium term, and long term), the utilization of physical approaches becomes imperative, particularly in short- and very-short-term horizons. This is due to the increasing significance of atmospheric dynamics, which have a more substantial impact on wind speed and power generation during these time frames [18].

Numerical weather prediction (NWP) models are mathematical models that provide information about the present and future state of the atmosphere and surface, including the ocean and land. These models typically have a forecast horizon of one to two weeks and are widely used in weather forecasting.

In a study presented in [19], the authors introduced a wind speed forecasting model that combines numerical weather prediction and historical measurements. The model utilizes multiple sources of past physical model outputs to enhance its forecasting accuracy. The developed model was applied to forecast wind speeds in a region near the U.S. Great Lakes. The results demonstrated an improvement in the root mean squared error of the proposed model, indicating its effectiveness in wind speed prediction.

Short-term wind forecasting using numerical weather prediction (NWP) models can be prone to significant errors due to its reliance on initial conditions. These models are slowly updated and may lag behind actual changes, leading to inaccuracies in short-term wind forecasts [20].

While AI-based methods, statistical methods, and hybrid models have been widely utilized for day-ahead wind speed forecasting, they may not be suitable for applications requiring high accuracy, such as operational control of a microgrid. The unpredictable nature of wind behavior and its direct correlation with physical indicators make it challenging for proposed models to achieve the desired level of accuracy in such scenarios. Additionally, NWP models have been employed for wind speed prediction; however, these models only consider current physical conditions and do not learn from past wind speed values or unexpected changes, limiting their predictive capabilities.

This paper aims to contribute to the advancement of knowledge in the field of wind speed forecasting, specifically for applications such as the design and planning of renewable energy systems. The proposed approach introduces a novel hybrid model that combines the Weibull distribution, long short-term memory (LSTM), and numerical weather prediction (NWP) models. The objective is to reduce the error associated with wind speed prediction by incorporating the distribution probability of historical wind speed data and considering the physical characteristics of the area.

The main contributions of this study, with respect to the prior literature, are as follows:

- Proposal of a hybrid model that overcomes the limitations of single statistical approaches. The LSTM method, which offers advantages over conventional feed-forward neural networks, is utilized in the proposed model;
- Introduction of a Weibull distribution of wind speed to capture the stochastic nature of wind behavior. By combining the probability distribution of wind speed with the LSTM model, the integrated model achieves a lower error compared to using a single LSTM model or a seasonal autoregressive integrated moving average (SARIMA) model with exogenous variables;
- Development of a hybrid model that integrates the results of the NWP model with AI models to enhance short-term forecasting accuracy (24–72 h). This hybrid model achieves minimal error and demonstrates the benefits of combining physical and AI-based approaches.
The remainder of this paper is organized into three main sections. Section 2 represents the related methodology of Weibull distribution and development of the LSTM model. Section 3 describes the results, including the comparison between each model and the final hybrid model. Finally, a conclusion providing a summary of the research and suggestions for future works is discussed in Section 4.

2. Methodology

This section provides an overview of the various forecasting models used in the study, as well as the proposed hybrid approach. The framework of the study is depicted in Figure 1, illustrating the flowchart that consists of three main sections.

The first section focuses on feature selection and data preprocessing. This includes tasks such as feature scaling, outlier detection, and handling missing values. Additionally, a grid search technique is utilized to identify the optimal parameters for the statistical model.

In the second section, the preprocessed data are used to train the developed models. Here, the hybrid model is constructed by incorporating the Weibull distribution output as
one of the input features for the LSTM model. Furthermore, the numerical weather prediction (NWP) data are extracted from the NWP model and utilized as inputs for the integrated LSTM–Weibull model, resulting in the final hybrid model.

The last section involves evaluating and comparing the accuracy of each model to determine the most suitable one. Hyperparameter optimization for the LSTM model is also performed in this section to fine-tune its performance.

2.1. LSTM

The recurrent neural network (RNN) is a deep learning algorithm commonly used for sequential data analysis, including time series data [21]. RNNs possess a unique feature called short-term memory, which is achieved through feedback connections within the network [22]. However, in practical applications, RNNs face challenges in capturing long-term dependencies in the data [23]. To overcome this limitation, the long short-term memory (LSTM) was developed as a specialized type of RNN [24,25].

The LSTM is a type of RNN proposed by Hochreiter and Schmidhuber in 1997 to deal with long-term dependencies by upgrading the remembering capacity of a simple recurrent cell [26].

An LSTM cell, in contrast to a simple RNN cell that consists of a single tanh layer [27]—which is a type of activation function commonly used in neural networks that squashes the input values to a range between −1 and 1—is composed of multiple layers, as illustrated in Figure 2. The initial layer is known as the forget layer, which determines whether the incoming information should be retained or discarded using an activation function. The name “forget layer” reflects its primary function of regulating the retention or deletion of prior information as new data are processed sequentially. Typically, the activation function used is a sigmoid function, which produces a value between 0 and 1 based on the input. A value of 1 indicates that the input can be added to the cell state, while a value of 0 signifies that the input should be forgotten or disregarded. \( f_t \) is the output of the forget layer, and it is determined using the below equation [28]:

\[
f_t = \varphi(w_f[y_{t-1}, x_t] + b_f)
\]

where \( \varphi \) is activation function, \( y_{t-1} \) is the output of the previous module, \( x_t \) is input at time \( t \) and \( b_f \) and \( w_f \) are bias and weight, respectively. This equation captures the role of the forget layer in LSTM, influencing the extent to which the cell retains or discards prior information for current predictions.

In the second step, the update of new values \( l_t \), using Equation (2), and a vector of new information \( g_t \), as shown in Equation (3), are created to add to the cell state by employing a sigmoid and tanh functions, respectively [28]:

\[
l_t = \varphi(w_i[y_{t-1}, x_t] + b_i)
\]

\[
g_t = \tanh(w_g[y_{t-1}, x_t] + b_g)
\]

Subsequently, in the third step, a new cell state \( g_t \) is expressed as the sum of the previous cell state multiplied by the first step results, and the multiplication of the \( l_t \) and \( g_t \) is shown in notational form as in the below equation [28]:

\[
g_t = f_t \times g_{t-1} + l_t \times g_t
\]

In the final step, by employing a sigmoid function, the cell decides what part of the cell state should be the cell’s output and input to the next cell, and by using a tanh function, it regenerates the values between −1 and 1 (Equations (5) and (6) [28]):

\[
\vartheta_t = \varphi(w_{\vartheta}[y_{t-1}, x_t] + b_{\vartheta})
\]

\[
y_t = \vartheta_t \times \tanh(g_t)
\]

where \( \vartheta_t \) is the portion of the cell’s state that is transmitted as the output.
Figure 2. Schematic design of an LSTM module.

The presence of the four layers within each cell of an LSTM model makes it a suitable algorithm to be evaluated for handling the unpredictable nature of wind speed.

2.2. Weibull Distribution

Wind speed can be expressed in time series, and the variation of the speed can be described using a probability distribution function (PDF). For many years, the Weibull distribution has been used to fit wind speed data, and it is an explicitly proper fit to average wind speed data [29].

The Weibull PDF can be described with Equation (7) [30]:

\[
F(v) = \left( \frac{k}{c} \right)^{k-1} \exp[-\left( \frac{v}{c} \right)^{k-1}]
\] (7)

where \( F(v) \) is the probability of occurrence of wind speed \( (v) \), and \( k \) is the Weibull shape parameter that is calculated based on the standard deviation \( (\sigma) \) and the average \( (\bar{v}) \) of the wind speed data using Equation (8) [31]:

\[
k = \left( \frac{\sigma}{\bar{v}} \right)^{-1.086}
\] (8)

And \( c \) is the Weibull scale parameter that is given as follows [29]:

\[
c = \frac{\bar{v}}{\Gamma\left(1 + \frac{1}{k}\right)}
\] (9)

where \( \Gamma \) is the gamma function.

2.3. SARIMAX

SARIMAX, which stands for Seasonal Autoregressive Integrated Moving Average with Exogenous Factors, is a statistical model commonly employed for time series prediction, specifically when there is seasonality present. It extends the SARIMA model by incorporating additional exogenous factors or predictors to further reduce the forecasting error. With considering \( y_t \) as the wind speed in time step \( t \), SARIMAX can be modeled as below [7,32–34]:

\[
\begin{align*}
\varphi & g_{t-1} + I_t \times \tanh \\
\varphi & y_{t-1} \times \tanh \\
\varphi & h_t
\end{align*}
\]
\[
\varphi_p(B) \Delta_p(B^s)(1 - B)^d(1 - B^s)^b y_t = \gamma_q(B) \delta_Q(B^s) \epsilon_t
\]  
(10)

where \( B \) is a lag operator that is responsible for back shifting, and \( \varphi_p(B) \) and \( \Delta_p(B^s) \) are non-seasonal and seasonal autoregressive operators of order \( p \) and \( P \), respectively. \( \gamma_q(B) \) and \( \delta_Q(B^s) \) are non-seasonal and seasonal moving average functions of order \( q \) and \( Q \), respectively. \( (1 - B)^d \) and \( (1 - B^s)^b \) are non-seasonal and seasonal differencing operators. \( s \) is the order of the seasonal component, and \( \epsilon_t \) is the residual error that captures the difference between the observed values and the values predicted by the model.

\[
\varphi_p(B) = 1 - \varphi_1(B) - \varphi_2(B^s) - \cdots - \varphi_p(B^p)
\]  
(11)

\[
\Delta_p(B^s) = 1 - \varphi_1(B^s) - \varphi_2(B^{2s}) - \cdots - \varphi_p(B^{ps})
\]  
(12)

\[
\gamma_q(B) = 1 - \gamma_1(B) - \gamma_2(B^2) - \cdots - \delta_q(B^q)
\]  
(13)

\[
\delta_Q(B^s) = 1 - \delta_1(B^s) - \delta_2(B^{2s}) - \cdots - \delta_Q(B^{qs})
\]  
(14)

Here, \( p, d, \) and \( q \) are integer parameters to show the delay order of non-seasonal autoregressive, differencing, and moving average terms, respectively, while \( P, D, \) and \( Q \) are integer parameters for indicating the delay order of seasonal autoregressive, differencing, and moving average terms, respectively. An optimum set of these parameters could be specified for the model as inputs using different criteria for parameter selection such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), or Hannan–Quinn information criterion (HQIC) methods. Furthermore, the seasonal length of the model should be estimated using the decomposition of the training data. Afterward, the model is applied to forecast the future wind speed. The forecast horizon has a direct impact on the accuracy of prediction. An increase in the length of the horizon results in a reduction in accuracy [7].

2.4. NWP Model

In this study, the NWP data were obtained from a model created by professionals with expertise in the field. The NWP model is a mathematical representation that characterizes the present and future state of the atmosphere and surface conditions, encompassing factors like temperature, pressure, humidity, and wind speed. It is formulated using established physical principles and numerical algorithms to simulate atmospheric behavior.

Although specific information regarding the development of the NWP model is not provided in this context, it is typically devised and upheld by meteorological agencies, research institutions, and weather forecasting centers. In this research, the NWP data were extracted from the NWP model that was developed by the “Centre for Solar Energy and Hydrogen Research (ZSW)” in Stuttgart, Germany.

2.5. Hybrid Model

In this study, the proposed hybrid model is built upon the foundational structure of the LSTM model. To integrate an additional model, referred to as Model X, with the LSTM model, the outputs of Model X are scaled and combined with other predictors and target variables. This augmented dataset is then used as input for the LSTM model.

By integrating Model X with the LSTM model, the overall input dimension of the neural network is increased by one, effectively adding an additional input parameter. This integration allows for the incorporation of additional information from Model X into the LSTM model, potentially enhancing the predictive capabilities of the hybrid model.
2.6. Preprocessing and Evaluation Metrics

Due to the use of several predictor variables, such as humidity, temperature, and air pressure, along with wind speed as the dependent variable, feature scaling is necessary to eliminate the issues associated with dimensionality caused by a dissimilar range of values. The min–max scaler method is used to scale the data into a similar range:

\[ X'_n = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  

(15)

To find the outliers in the dataset based on an extreme outlier detection procedure, the minimum and maximum bounds were calculated based on Equations (16) and (17), respectively.

\[ Q1 - 3(IQR) \]  

(16)

where Q1 is the lower quartile that shows the number that is more than 25 percent of the data, and IQR is the interquartile range.

\[ Q3 + 3(IQR) \]  

(17)

where Q3 is the upper quartile that shows the number that is more than 75 percent of the data.

To assess the forecasting models’ performances, root mean squared error (RMSE), mean absolute error (MAE), and mean squared logarithmic error (MSLE) are employed to determine the goodness of fit. RMSE evaluates the error using Equation (18):

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\tilde{y}_i - y_i)^2}{n}} \]  

(18)

where \( y \) is the observed value, and \( \tilde{y} \) is the predicted value.

And MAE is calculated using Equation (19)

\[ \text{MAE} = \frac{\sum_{i=1}^{n} |\tilde{y}_i - y_i|}{n} \]  

(19)

Due to the wind speed, as the target variable is distributed based on Weibull distribution, and the considerable difference between the minimum and maximum value in wind speed data, MSLE could be a proper metric to evaluate the error of a model. It could be calculated using Equation (20):

\[ \text{MSLE} = \frac{1}{n} \sum_{i=1}^{n} (\log(y_i + 1) - \log(\tilde{y}_i + 1))^2 \]  

(20)

3. Case Study and Data Characteristics

In this research, Montreal, the second-most populous city in Canada, is considered as the case study. Montreal is located in the southern part of the province of Quebec, Canada with latitude and longitude coordinates of 45 N and –73 E degrees, respectively [35]. Montreal’s hourly resolved data of temperature and humidity were obtained from NASA’s prediction of worldwide energy resources website [36]. The data for predictors were collected from Jan 2020 until Jan 2021 for training and test purposes. Figure 3 shows the variation and trend of the independent variables. Although the ascending and descending trend from the beginning to the end of the year is noticeable for temperature, no notable trend is detected in relative humidity. However, a lower fluctuation range at the beginning of the year (winter) compared to the middle of the year (summer) is perceptible for relative humidity.

Furthermore, the target variable, Montreal’s wind speed (in m/s) at 50 m in height, is also collected from [36] in an hourly resolution from January 2020 to January 2021 for...
training and testing purposes. To evaluate the behavior of the target variable further, the additive decomposition of the wind speed data is plotted (Figure 4). As it is demonstrated in Figure 4, although there is no notable trend, the seasonality graph shows daily fluctuations (ascending and then descending during a day). However, as the residual graph that shows the error of fitting this seasonality on the real wind speed data is noticeable, this seasonality can be seen as not being strong.

Figure 3. Predictors' overall trend and seasonality in the training set.

Figure 4. Additive decomposition of training data.

4. Results
4.1. Implementation

LSTM, SARIMAX, and the proposed hybrid methods were developed in the Python programming language (Version 3.7.9). To generalize the results of the testing of the
developed models for the whole year, three different test sets from summer (the last 2 days of July 2020), fall (the last two days of October 2020), and winter (the last two days of December 2020) were selected as the representative of different seasons. For summer, the model trained with the data from January 2020 until 29 July 2020, while for Fall and Winter, the training set included data from January 2020 until 29 October 2020 and 29 December 2020, respectively. A few missing values were found in the training sets, and they have all been replaced by the average of the previous and next values. Also, the outlier detection procedure was implemented using boxplot visualization and by calculating quartiles based on the formula explained in the methodology section. The results showed no outlier in the training sets.

To scale up the predictors and target variable into a unique scale, the MinMaxScaler method from preprocessing sub-package of the Sklearn library (Version 0.24.2) was used. All features were scaled into the range between 0 and 1 before feeding to the neural network model.

To form the LSTM layers, the Keras library (Version 2.8.0) with TensorFlow (Version 2.8.0) backend was used. Furthermore, grid search optimization was applied to find the optimum hyperparameters (Table 1). The hyperparameter optimization shows that the combination of Adam optimizer and a batch size of 32 with 150 epochs results in with only a minor loss in the training stage. A further increase in the number of epochs results in further reduction in errors with the training data, as shown in Figure 5; however, above 150 epochs, the overfitting tends to cause a reduction in accuracy of the model in forecasting the test dataset. As explained in the previous section, the trained model was then tested on the first two days of July 2020. The training and test datasets are of the same resolution.

Table 1. Hyper parameters of the LSTM model and parameter selection results.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values/Types</th>
<th>Hyper Parameter Optimization Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. hidden layers</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>No. neurons per hidden layer</td>
<td>60</td>
<td>-</td>
</tr>
<tr>
<td>Activation function</td>
<td>Sigmoid</td>
<td>-</td>
</tr>
<tr>
<td>Optimizer types</td>
<td>{Adam, RMSprop}</td>
<td>Adam</td>
</tr>
<tr>
<td>Batch size</td>
<td>{1,32,64}</td>
<td>32</td>
</tr>
<tr>
<td>No. of epochs</td>
<td>{50,80,100,110,125,150}</td>
<td>150</td>
</tr>
</tbody>
</table>

Figure 5. Model convergence plot for single LSTM model.
The Weibull model was developed by creating a function to generate the wind speed distribution. The Weibull distribution of the wind speed for the year 2020 was calculated in the Python environment by creating a Weibull function using $c$, $k$, and $\Gamma$ parameters that have been explained in the methodology section. The histogram graph in Figure 6 shows the data distribution in the range of 0–20 m/s. Also, the Weibull probability feature was created using the Stats package of the Scipy library (Version 1.7.0).

The parameter selection of the SARIMAX model was made by applying the Autoarima package from the Pmdarima library (Version 1.8.2) and a grid search through 42 different combinations of the $(p, d, q)(P, D, Q, s)$ parameters. The last two months of wind speed historical data of the first half of the year 2020 were used for training the Autoarima for parameter selection. The results (Table 2) show that the combination of $(2,0,1)(2,1,0,24)$ yields the minimum AIC and was selected as the optimum set of parameters for the SARIMAX model. All the other combinations that are not mentioned in Table 2 led to AIC equal to infinity.

![Figure 6. Weibull distribution of the historical wind speed.](image)

**Table 2.** Report on the grid search for SARIMAX parameter selection.

<table>
<thead>
<tr>
<th>Combination</th>
<th>AIC</th>
<th>Combination</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0,0,0)(0,1,0,24)$</td>
<td>11,271</td>
<td>$(1,0,1)(2,1,0,24)$</td>
<td>4044</td>
</tr>
<tr>
<td>$(1,0,0)(1,1,0,24)$</td>
<td>4955</td>
<td>$(2,0,1)(2,1,0,24)$</td>
<td>3959</td>
</tr>
<tr>
<td>$(1,0,0)(0,1,0,24)$</td>
<td>5548</td>
<td>$(3,0,1)(2,1,0,24)$</td>
<td>3961</td>
</tr>
<tr>
<td>$(1,0,0)(2,1,0,24)$</td>
<td>4709</td>
<td>$(2,0,2)(2,1,0,24)$</td>
<td>4221</td>
</tr>
<tr>
<td>$(0,0,0)(2,1,0,24)$</td>
<td>10,553</td>
<td>$(1,0,0)(2,1,0,24)$</td>
<td>4707</td>
</tr>
<tr>
<td>$(2,0,0)(2,1,0,24)$</td>
<td>3998</td>
<td>$(1,0,2)(2,1,0,24)$</td>
<td>3972</td>
</tr>
<tr>
<td>$(2,0,0)(1,1,0,24)$</td>
<td>4211</td>
<td>$(3,0,0)(2,1,0,24)$</td>
<td>3962</td>
</tr>
<tr>
<td>$(3,0,0)(2,1,0,24)$</td>
<td>3964</td>
<td>$(3,0,2)(2,1,0,24)$</td>
<td>3963</td>
</tr>
<tr>
<td>$(3,0,0)(1,1,0,24)$</td>
<td>4175</td>
<td>$(4,0,1)(2,1,0,24)$</td>
<td>3963</td>
</tr>
<tr>
<td>$(5,0,0)(2,1,0,24)$</td>
<td>3965</td>
<td>$(3,0,1)(1,1,0,24)$</td>
<td>4176</td>
</tr>
</tbody>
</table>

The results consist of details of the selected combinations, including the AIC, BIC, and HQIC, which are reported in Table 3.
Table 3. Selected combination information.

<table>
<thead>
<tr>
<th>Parameter/Metric</th>
<th>Value/Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimum non-seasonal orders</td>
<td>(2,0,1)</td>
</tr>
<tr>
<td>Optimum seasonal orders</td>
<td>(2,1,0,24)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>2160</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-4270.623</td>
</tr>
<tr>
<td>AIC</td>
<td>3959.516</td>
</tr>
<tr>
<td>BIC</td>
<td>4004.849</td>
</tr>
<tr>
<td>HQIC</td>
<td>3976.106</td>
</tr>
<tr>
<td>Covariance type</td>
<td>Outer Product of Gradients (OPG)</td>
</tr>
</tbody>
</table>

4.2. Discussion

Figure 7 shows the forecasting results of all three models and the hybrid models’ results for the last two days of July, October, and December 2020. At a glance, the results using a single LSTM model do not show proper wind speed forecasting, especially in peak hours that are way over or under actual values. By applying the SARIMAX model, although the mean value of the forecasted wind speeds is nearer to the mean value of the actual wind speed compared with the single LSTM model, it has not captured the fluctuations, peaks, and trends decently. While using the NWP model that resulted in a significant error, especially in winter, applying the proposed integrated model can considerably reduce this error. Furthermore, the seasonality and trend issues seem to be fixed for the whole prediction horizon in different seasons. However, the accuracy is not high in the two major peaks. A quick comparison between the result of the proposed hybrid model and the result of the other models reveals the hybrid model’s ability to better integrate the fluctuations and trends.

To evaluate and compare the models precisely, the RMSE, MAE, and MSLE of each model’s results are calculated based on what was explained in the methodology section. The results are reported in Table 4. The LSTM model results show a 2.21–3.16 root mean squared error in different seasons that depicts the LSTM model’s inability to accurately predict using the three meteorological historical data (temperature and humidity) as the independent variables. Although with the SARIMAX model, the RMSE and MSLE are improved in fall, the error is still high. As explained in the methodology section, the LSTM model with different layers in its cells can deal with unexpected behavior of data. Therefore, a proper feature should be added to the LSTM model for better training. Integrating the probability distribution of the wind speed with the LSTM model and using it as an input feature could be one of the alternatives to boost the LSTM ability. The results in Table 4 show that the proposed integrated LSTM–Weibull model can reduce the RMSE of the single LSTM model in forecasting winter, summer, and fall representative days by about 13, 39, and 31 percent, respectively. These error reductions show that adding a proper feature, such as the Weibull probability of the wind speed, can help LSTM accurately forecast the future. However, in case of any unexpected wind behavior that has not happened before (and it is normal in climatic situations), even the integrated model could lead to a considerable error. Therefore, to solve this challenge, hybridizing the results of the NWP model predictions with the proposed integrated model could be helpful in capturing the unexpected behavior of the wind that was not recorded in the historical data. The single NWP prediction results also show high RMSE and MAE and even higher MSLE compared with other models, especially in winter and fall. However, by hybridizing the NWP model with the integrated model, the RMSE of the proposed model decreased 47%, 17%, and 32%, respectively, in summer, winter, and fall compared with the single LSTM model.

To consider the effect of the prediction horizon on the final accuracy, the prediction periods were extended to 168 h (one week) for all seasons instead of 48 h (two days). Since the look-back period is 48 h, it means that after forecasting the first 48 h into the future,
the next hours will be predicted based on the prior predictions. Therefore, the accuracy of the model could be lower with increasing the prediction horizon. The result of the prediction horizon extension is shown in Figure 8 for the hybrid model. It is evident that the hybrid model acts less and less accurately when increasing the prediction period except for fall, which still can predict the third day (until 72 h) correctly, and that could be because of the fewer fluctuations on the third day.

![Figure 7. 48 h forecasting results in different seasons.](image)

Table 4. Evaluation metrics of all the models for wind speed forecasting.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>MSLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>July (Summer)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>2.21</td>
<td>1.86</td>
<td>0.395</td>
</tr>
<tr>
<td>SARIMAX</td>
<td>2.64</td>
<td>2.24</td>
<td>0.234</td>
</tr>
<tr>
<td>NWP</td>
<td>2.02</td>
<td>1.70</td>
<td>0.227</td>
</tr>
<tr>
<td>Integrated LSTM–Weibull</td>
<td>1.35</td>
<td>1.12</td>
<td>0.071</td>
</tr>
<tr>
<td>Hybrid LSTM–Weibull–NWP</td>
<td>1.18</td>
<td>0.95</td>
<td>0.066</td>
</tr>
<tr>
<td>December (Winter)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM</td>
<td>2.14</td>
<td>1.63</td>
<td>0.106</td>
</tr>
<tr>
<td>SARIMAX</td>
<td>2.81</td>
<td>2.3</td>
<td>0.155</td>
</tr>
<tr>
<td>NWP</td>
<td>5.58</td>
<td>4.90</td>
<td>0.849</td>
</tr>
<tr>
<td>Integrated LSTM–Weibull</td>
<td>1.87</td>
<td>1.55</td>
<td>0.110</td>
</tr>
<tr>
<td>Season</td>
<td>Hybrid LSTM–Weibull–NWP Mean</td>
<td>LSTM Mean</td>
<td>NWP Mean</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>October (Fall)</td>
<td>1.78</td>
<td>1.50</td>
<td>0.078</td>
</tr>
<tr>
<td>LSTM</td>
<td>3.16</td>
<td>2.58</td>
<td>0.272</td>
</tr>
<tr>
<td>SARIMAX</td>
<td>2.25</td>
<td>2.73</td>
<td>0.294</td>
</tr>
<tr>
<td>NWP</td>
<td>4.14</td>
<td>3.11</td>
<td>0.799</td>
</tr>
<tr>
<td>Integrated LSTM–Weibull</td>
<td>2.18</td>
<td>1.60</td>
<td>0.100</td>
</tr>
<tr>
<td>Hybrid LSTM–Weibull–NWP</td>
<td>2.16</td>
<td>1.67</td>
<td>0.139</td>
</tr>
</tbody>
</table>

Figure 8. One week (168 h) forecasting results of the hybrid model in different seasons.

5. Validation

Since training the model with different types of historical data and parameters such as learning rate could lead to different results [37], from what was shown in Table 4 and from significant result changes from changing the forecasting horizon, validating the results of this study with the results of the other research could not be insightful. However, based on the literature [38], the mean absolute percentage error for wind forecasting ranged between 25% to 40%. Furthermore, based on similar research [4] that forecasted wind speed in four different sites in China, the RMSE of their proposed hybrid model for daily wind forecast ranged between 1.6–1.8. Comparing this result with the output of the hybrid model presented in this research (Table 4) for a two-days-ahead forecast shows acceptable prediction accuracy.

6. Conclusions

Wind power stands as a pivotal source of clean energy, poised to synergize with other renewables for fostering a robust future grid. Nonetheless, the intermittent nature of wind resources mandates accurate wind speed or wind power predictions to optimize grid control and unit dispatch. Addressing the challenges posed by the volatile wind speed patterns and the complexities of deciphering genuine daily trends and seasonality in historical data, this study sought to pioneer an innovative hybrid wind speed forecasting model.
This model harnesses the prowess of deep learning, probability distributions, and numeric weather prediction techniques to minimize forecasting errors.

Our findings illuminate the limitations of the LSTM model, which, despite its memory mechanisms, falters in precise predictions—particularly during abrupt surges or unforeseen shifts. Our initial innovation, fusing Weibull distribution probabilities with a singular LSTM model, resulted in remarkable error reduction. Specifically, the hybridization yielded an average RMSE reduction of approximately 28% across three diverse prediction horizons throughout the year.

To account for unforeseen wind behavior unrepresented in historical data, we augmented our approach by hybridizing results from the numeric weather prediction model into the LSTM–Weibull integrated framework. The outcomes showcased the final hybridized model’s prowess in diminishing the average RMSE of the solo LSTM predictions by around 32%, especially when confronted with fluctuations occurring in central peaks. This hybrid model emerges as a promising avenue for curbing wind speed forecasting errors, thus offering a potential catalyst for robust management and control of renewable energy systems.

Looking ahead, future research should explore synergizing the proposed model with statistical approaches like SARIMAX or ARIMA to amplify wind speed forecasting precision. Such a comprehensive amalgamation would capitalize on the strengths of distinct models, effectively addressing individual method limitations.

In summation, our proposed hybrid model paves the way toward heightened wind speed forecasting accuracy—a critical stride toward efficient renewable energy system management and control. As the renewable energy landscape evolves, the fusion of pioneering methodologies promises an increasingly sustainable and efficient energy future.

Author Contributions: Conceptualization, N.S.; Methodology, N.S.; Software, N.S.; Validation, N.S.; Formal analysis, N.S.; Investigation, N.S., F.N., R.P.M. and U.E.; Resources, N.S. and A.K.; Data curation, N.S. and R.P.M.; Writing—original draft, N.S.; Writing—review & editing, F.N., R.P.M., P.M., A.K. and U.E.; Visualization, N.S.; Supervision, F.N. and U.E.; Project administration, F.N. and U.E.; Funding acquisition, F.N. and U.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [NSERC Discovery grant] grant number [RGPIN-2016-06727] and also the [Canada Excellence Research Chair in Smart, Sustainable and Resilient Communities and Cities] funded by [Tri-Agency Institutional Program Secretariat].

Data Availability Statement: Not applicable.

Acknowledgments: We would like to acknowledge the financial support of the NSERC discovery grant program; the Gina Cody School of Engineering Faculty Research Support; and the Canada Excellence Research Chair in Smart, Sustainable and Resilient Communities and Cities funded by the Tri-Agency Institutional Program Secretariat. We also extend our thanks to the Centre for Solar Energy and Hydrogen Research (ZSW) for generously providing the numerical weather prediction data.

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

References


**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.