Received: 25 January 2022 | Revised: 4 September 2022 | Accepted: 11 November 2022

DOI: 10.1002/ep.14039

ORIGINAL RESEARCH

Sustainable Energy

ENVIRONMENTAL PROGRESS & SUSTAINABLE ENERGY

The impact of input data resolution on neural network forecasting models for wind and photovoltaic energy generation using time series data

Mutaz AlShafeey | **Csaba Csaki**

Institute of Informatics, Corvinus University of Budapest, Budapest, Hungary

Correspondence
Mutaz AlShafeey, Institute of Informatics, Corvinus University of Budapest, Budapest, Hungary.
Email: mutaz.alshafeey@uni-corvinus.hu

Funding information
3Comm Hungary; National Research, Development and Innovation Fund of Hungary, Grant/Award Number: TKP2020-NKA-02; SolarEdge

Abstract

Accurate photovoltaic (PV) and wind energy forecasting are crucial for grid stability and energy security. There are various modeling techniques and methods to design forecasting models, each leading to different accuracy. In this research, datasets were collected from a 546 kWp grid-connected PV farm and a 2 MW wind turbine for one full year. These data were used to train and test artificial neural network models to forecast day-ahead PV and wind energy utilizing time-series input data with 15-, 30-, and 60-min resolutions. The models were able to forecast the PV energy accurately, while the same models trained for wind showed poor performance. Higher input data resolutions lead to slightly better forecasting performance for the PV farm. Utilizing data with higher resolution can improve the forecast by 1%–5%. While for wind energy forecasting, the resolution has very minor effects, although the 30-min resolution shows a slightly better forecasting performance.

KEYWORDS

ANN forecasting models, day-ahead energy prediction, forecasting resolutions, PV energy forecasting, wind energy forecasting

1 | INTRODUCTION

The reliance on fossil fuels and their derivatives for energy generation might have serious consequences on the environment. Issues such as climate change, greenhouse effect, and deforestation can increasingly be linked to fossil fuel dependency.[1] Moreover, fossil fuel is a depleted source that is not distributed evenly around the globe. Hence, inequality in the distribution of energy consumption and reserves is another problem for current energy systems.[2]

Fossil fuel dependency problems can be reduced by renewable sources. Energy generation from solar, wind, tidal wave, or biomass can offer a reliable and cost-effective solution. These renewable resources are expected to have significant advantages over their conventional counterparts.[3]

Since photovoltaic (PV) equipment can be easily installed almost everywhere and operates efficiently in different geographical regions with low maintenance required, solar energy is considered to be an effective environmentally friendly technology for energy production.[4] Another growing trend in renewable energy generation is the utilization of wind resources. Wind technologies offer reliable, eco-friendly, simple, and low-maintenance methods for energy generation.[5]

Despite all the attractive advantages of utilizing solar and wind technologies, some major challenges limit their wider applicability and impact national strategies and policies of renewables. The fluctuations in wind and solar resources create ambiguities in the produced energy.[6] Uncertainties in energy production affect energy security, economic performance, and most importantly, grid stability (in case of grid-connected solutions). These issues are further exposed when considering them in a regulatory context, especially from an international perspective.[7]

Most regions and jurisdictions have their own regulations regarding production, trade, and distribution of electricity—such as the "Clean Energy for All" package of the European Union that affects 28 member states.[8] National policies also need to address market realities and have to keep up with the rapid change induced by the integration of an increasing ratio of renewable-based power.[9]

The core of most rules focuses on scheduling processes and related services such as day-ahead and intraday scheduling and balancing. Consequently, the accuracy of production prediction has become crucial. Policy handling of this complex setting, such as incentives or pricing does impact investment calculations for the long run and production control for the short term (e.g., the case of Germany[10]). Hence, effective ways of integrating these intermittent resources with electricity grids are needed. Renewable energy integration issues influence the formulation of regulations regarding renewable energy production, trade, and distribution.

Renewable regulations have regularly been augmented and modified to keep up with the need to contain market abuses and keep grid operations safe. Some of the integrating methods use energy storage systems to stabilize the power. Yet, using storage units is impractical for large applications. In addition, storage units like batteries have limited usage cycles and it has to be replaced after a certain time adding extra costs.[11]

Another method for better integration is forecasting energy production.[12] Accurate energy forecasting models do not only provide value through reduced imbalance penalties (incurred due to the difference between the scheduled and actually delivered energy) but also lead to increased competitiveness by providing advanced knowledge in real-time energy market trading.

Since renewables are erratic, difficult to predict, and challenging to integrate with the existing systems, implementing circumspection and improving energy policy should be vital parts of this energy transition.

Given that forecasting models are able to address the complex patterns of energy production based on renewable resources and can also handle the typical uncertainties of energy demand, sustainable energy policies and scenarios should consider such models.[13] Low error production predictions improve the construction, operations, and maintenance planning of energy projects.[14] Forecasting of renewable energy generation is also a vital part and the base to plan, design, and manage the energy supply policy.[15] This means that good forecasting tools would positively impact both the cost and integration of wind and solar farms.

Many renewable energy forecasting methods have been used over the years. Physical models based on numerical weather predictions (NWPs), statistical and probabilistic models, and artificial intelligence (AI) models based on machine learning (ML) are among the main utilized techniques.[16]

Moreover, each forecasting method has different techniques that can be used to forecast renewable energy production, especially for wind and solar. For instance, statistical Seasonal Autoregressive Integrated Moving Average (SARIMA) model was presented by Vagropoulos et al.[17] to predict the performance of a grid-connected PV farm. But statistical methods might not be the best approach to support decision-makers as these models are very complex and the prediction accuracy decreases for longer horizons.[18]

To offer an alternative to statistical prediction approaches in order to overcome their weaknesses, machine learning (ML) forecasting models based on deep learning have been proposed. ML has several modeling algorithms like supervised, unsupervised, and meta-learning algorithms, each used for specific learning tasks.[19] Artificial neural networks (ANN) is one of the supervised ML algorithms, which can be used to solve complex nonstationary and nonlinear problems.

ANN can be defined as a set of connected units called artificial neurons arranged into structural layers. The connection network between the neurons is similar to the synapses in a biological brain. Each neuron can receive, transmit, and process signals from and to other neurons connected to it (that are usually located in a different layer).

Some ANN forecasting models show very good abilities in predicting solar and wind energy with minimum errors and lowered uncertainties compared to other ML algorithms.[16] Yet, forecasting the potential of wind and solar energy is not an easy task as many factors impact performance and forecasting accuracy.

Besides the modeling techniques and algorithms applied, the data forming the input to the models as well as the forecasting horizon and resolution might also affect the performance of these models.[20] Depending on the input data (i.e., explanatory variables) utilized by the forecasting models (including ANN), there are three main methods for building the forecasting models[21,22]: (1) in the structural method forecasting models utilize geographical and meteorological parameters (such as wind speed, ambient temperature, humidity, and so on); (2) in the time-series method only past power values are utilized by the models as inputs; (3) in the hybrid method both meteorological variables and past power values are utilized by the forecasting models.

In order to create accurate energy plans and stabilize energy infrastructures, many studies have developed renewable energy forecasting models based on time-series input. For instance, Reikard[23] ran experiments on six data sets at resolutions of 5, 15, 30, and 60 min using the global horizontal component of solar radiation to forecast PV energy utilizing different models. It was found that the ARIMA method has better abilities in capturing the diurnal cycle more effectively than other methods tested. Chang used an ANN with a radial basis function to forecast wind power.[24] Heinermann and Kramer[25] used the decision tree and the support vector machine techniques to build a heterogeneous ML model ensemble for predicting wind power. The suggested method shows better results than state-of-the-art machine learning methods.

According to the time horizon of the forecast, energy forecasting can be divided into four major types: very short-term (few seconds to

30 min), short-term (30 min–6 h), medium-term (6 h–1 day), and long-term forecasting (days, weeks, etc). It should be noted here, that time-scale classification of forecasting models in the literature is relatively vague.[26] Owing to simplicity and high accuracy, many existing works focus on very short-term or short-term energy forecasting.[27] Generally, regardless of the modeling technique or data utilization method used, forecasting accuracy is expected to decrease for longer horizons.[28]

Moreover, the input data resolution (frequency of the input variables per time unit) is also affecting the accuracy of the forecast.[29] Generally, past energy production data are collected with high sampling resolution, such as 10 min,[30] 15 min,[31] and 1 h.[32] When a longer prediction horizon is required, the original high-resolution data are usually averaged to build up low-resolution data.[33] Unfortunately, the process of averaging will lead to a lot of information losses as the rapid fluctuations in original high-resolution data will be neglected.[34]

While it has been established that different forecasting horizons lead to different accuracies, the impact of input data resolution could bear some clarification. Considering the above challenges and options, the research reported here developed ANN-based energy forecasting models for both PV and wind renewable energy technologies. The goal was to investigate and compare the performance of ANN time-series forecasting models of both PV and wind energy for 24-h (day-ahead) horizons under different input data resolutions. Testing was aimed at finding the input data resolution that leads to the best accuracy for a 24-h forecasting horizon, leading to improved day-ahead energy scheduling and for consideration in policy expectations concerning delivery scheduling.

Although there are claims stated that if the data resolution is higher, the model developed in any way will perform better, this article clarifies in detail the effects of utilizing different input data resolutions on several forecasting models' performance measures for both PV and wind farms. This article also clarifies that utilizing input data with higher data resolutions might not always lead to better forecasting accuracy compared with lower resolutions.

To present the findings, the paper is organized as follows: the article starts with an introduction, that includes a background review on renewable energy prediction models with a special focus on ANN. Section 2 then presents the model-building process along with data sources, model details, and indicators used in the comparison. This is followed by a presentation of the results and a discussion of the findings. The paper closes with conclusions, limitations, and directions for future research.

2 | DESIGN OF THE EXPERIMENT, DATA COLLECTION, MODELS, AND EVALUATION METHODS

To address the above objective, two sets of ANN time series forecasting models were designed, built (i.e., trained), and tested to forecast wind and PV out power for 24 h ahead, each set utilizing input data with resolutions of 15, 30, and 60 min. Once the six models were trained, their accuracy was then calculated. Subsequently, a comparative analysis was conducted to determine the best settings leading to the best performance.

The input to the time-series ANN models are past energy values, therefore, to train the ANN models, both actual PV and wind past energy values were collected covering a bit more than 13 months. Data collection started on May 1, 2019, and lasted till June 13, 2020. PV data were collected from a 546 kWp grid-connected solar farm located in Hungary. While wind data were collected from a 2 MW wind turbine located also in Hungary. All data were collected in 15-, 30-, and 60-min resolutions.

Figure 1 shows the overview of the methodology. The process starts by collecting the past generation data for the PV farm and wind turbine. The data are used in its original resolution as collected, thus, data were not averaged to build up lower resolutions. Then six ANN forecasting models were designed and trained: ANN models were built to forecast PV and wind energy both with 15-, 30-, and 60-min resolutions. The target horizon of the forecast is 24-h ahead. The outputs (forecasted values of PV and wind energy) were then stored. After 24-h delay, when the real generation values have become available (as the real production values are always lagging 24 h behind the forecasted ones), the performance of each model was calculated. The output data are used to update the historical records and then to continue the training of the models.

Figure 2 shows a simple diagram of ANN with *n* number of inputs and one output. The input variables arrive from the bottom (input layer) and can pass through the middle layer(s) to reach the succeeding ones, while the forecasted variable(s) (output) are at the top layer (the output layer). Neural networks may possibly include one or more hidden layers with hidden neurons (generally called nodes). The function of such nodes is to perform a nonlinear transformation on the input data entering that hidden neuron (as can be seen in Figure 3). This way hidden neurons receive inputs from the nodes of the previous layer and deliver the calculated result to nodes in the next layer. Neurons in a given layer may be fully or partially connected to each neuron in the next layer. Additionally, the error of the output is calculated and then utilized to tune the network.

The ANN structure where the direction of information flows from the bottom up to the top layers in one direction is known as multilayer feed-forward neural network (MLFFNN). Using the MLFFNN network structure has many advantages for this particular research context such as the ability to solve complex nonlinear problems or the ability to achieve good accuracy with smaller data sets. MLFFNN works well with big data and can provide quick results after training.[35,36] Therefore, regarding the prediction models reported here, a MLFFNN was designed.

In any given MLFFNN layer, each neuron combines its inputs using a weighted linear combination as shown in Equation 1 where $\boldsymbol{v_1},\boldsymbol{v_2},\ldots\ \boldsymbol{v_n}$ are that neurons' inputs coming from each neuron in the previous layer, $\boldsymbol{w_s}$ is the weighted sum, $\boldsymbol{w_1},\boldsymbol{w_2},\ldots\boldsymbol{w_n}$ are the weights corresponding to the inputs, and $\boldsymbol{b}$ is the bias:

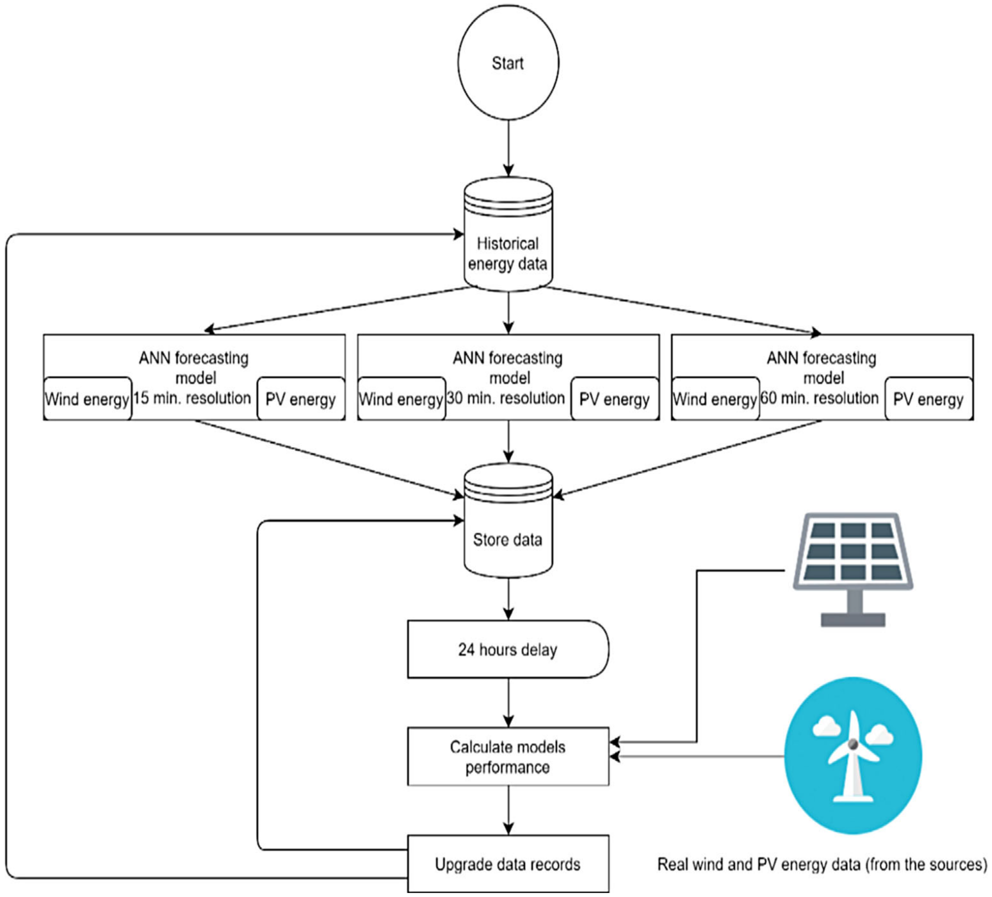

FIGURE 1 Flowchart providing a general overview of the methodology

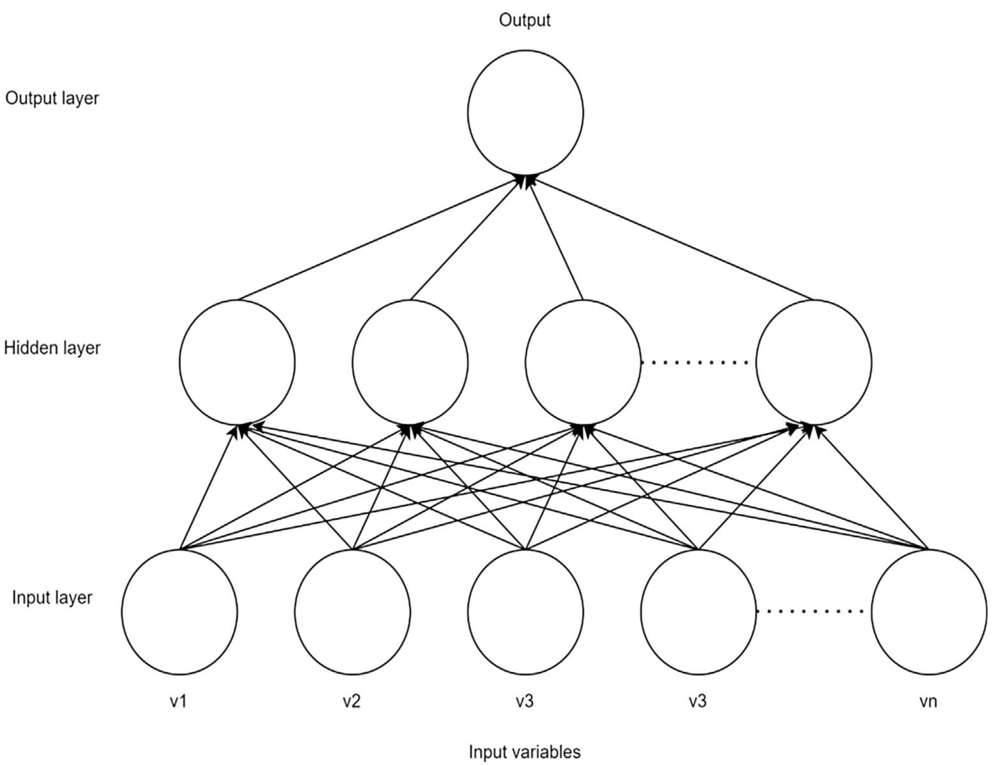

FIGURE 2 A neural network with *n* inputs and one output

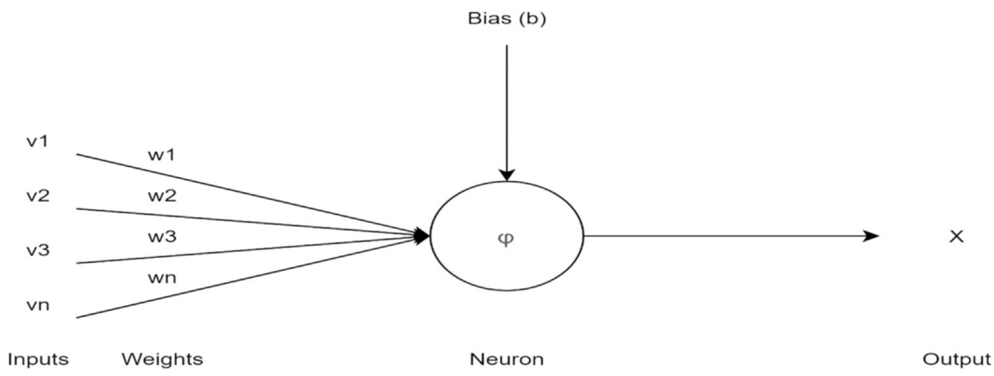

FIGURE 3 Flow of information in an artificial neuron

$$w_s = w_1 \times v_1 + w_2 \times v_2 + \dots + w_n \times v_n + b \tag{1}$$

Although in this experiment fully connected layers are used, one must note, that in case of non-fully connected structures, some weights might be set to zero constant (depending on some structuring principle). The weighted sum equation can be written with matrices as in Equation (2)[37]:

$$w_s = wv + b \tag{2}$$

where w and v are defined as:

$$w = [w_1 w_2 w_3 \dots\dots w_n] \text{ and } v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ . \\ . \\ . \\ v_n \end{bmatrix} \tag{3}$$

Then, a nonlinear transfer function modifies the results as in Equation (4), where X is the neurons' output:

$$X = \varphi(w_s) \tag{4}$$

Different transfer functions might be used, yet, the sigmoid function is among the most used[38] as can be seen in Equation (5):

$$\varphi(ws) = \frac{1}{1 + e^{-ws}} \tag{5}$$

Tansig transfer function was used instead of the sigmoid in case of negative values are found in the output as in Equation (6):

$$\varphi(ws) = \frac{1 - e^{-2ws}}{1 + e^{-2ws}} \tag{6}$$

The equations above illustrate the mathematical representation for a single neuron. Notice that each neuron has its specific set of weights and biases (see Equations 1 and 2). Initially, each (hidden and output) neuron's weights are set to random values. Then, training data are fed to the input layer of the ANN. The data then move bottom-up through the layers, getting modified and adjusted as illustrated in the equations, until it finally reaches the output layer significantly transformed. Then, as mentioned earlier, an error is calculated and used to train the network by modifying the weights and the biases. In this research the back-propagation algorithm applied is based on the minimization of the mean square error (MSE) between the real and the output data. In MLFFNN, the MSE is minimized in proportion to the input value(s) and the output value(s) as can be seen in Equation (7)[39]:

$$\text{Min(MSE)} = \min\left(\frac{1}{n} \times \sum_{i=1}^{n} (y_t - p_t)^2\right) \tag{7}$$

The changes in weights and biases are calculated according to Equations (8) and (9), respectively[39]:

$$\Delta w_n = \gamma(y_n - p_n) \tag{8}$$

$$\Delta b_n = \gamma(y_n - p_n) \tag{9}$$

where Δw_n and Δb_n is the change of the weight and bias, respectively, for the nth neuron, and γ is the learning rate. Consequently, the adjusted weight (w_{adjusted}) and bias (b_{adjusted}) can be calculated according to Equations (10) and (11) correspondingly:

$$w_{\text{adjusted}} = w + \Delta w \tag{10}$$

$$b_{\text{adjusted}} = b + \Delta b \tag{11}$$

One cycle of the above-mentioned process (when training data are fed to the input layer of the ANN and pass forward through the succeeding layers and then weights of each neuron are adjusted based on the MSE of the resulting output) is called an epoch. Such epoch loops will continue until the MSE reaches the lowest possible limit (generally when the MSE value does not change for several epochs) or when a given number of epochs is reached.

In this research, PV and wind forecasting models using fully connected MLFFNNs were built and tested with three different resolutions. This implies a differing number of input neurons for each resolution tested. The number of input neurons, therefore, are 96, 48, and 24 for the 15, 30, and 60 input resolutions, respectively, for both PV and wind. Another important parameter for ANN is the number of hidden neurons. Few hidden neurons might affect the ability of ANN

TABLE 1 ANN parameters

Parameter	Description	Value for each resolution 15	30	60
Number of inputs	Number of input data variables	96	48	24
Number of outputs	Number of output forecasted variables	1	1	1
Number of hidden neurons	Number of hidden neurons	32	16	8
Maximum Epochs	Max. number of training iterations before training is stopped	1000	1000	1000
Maximum training time	Max. time before training is stopped	∞	∞	∞
Performance Goal	The min. target value of MSE	0	0	0

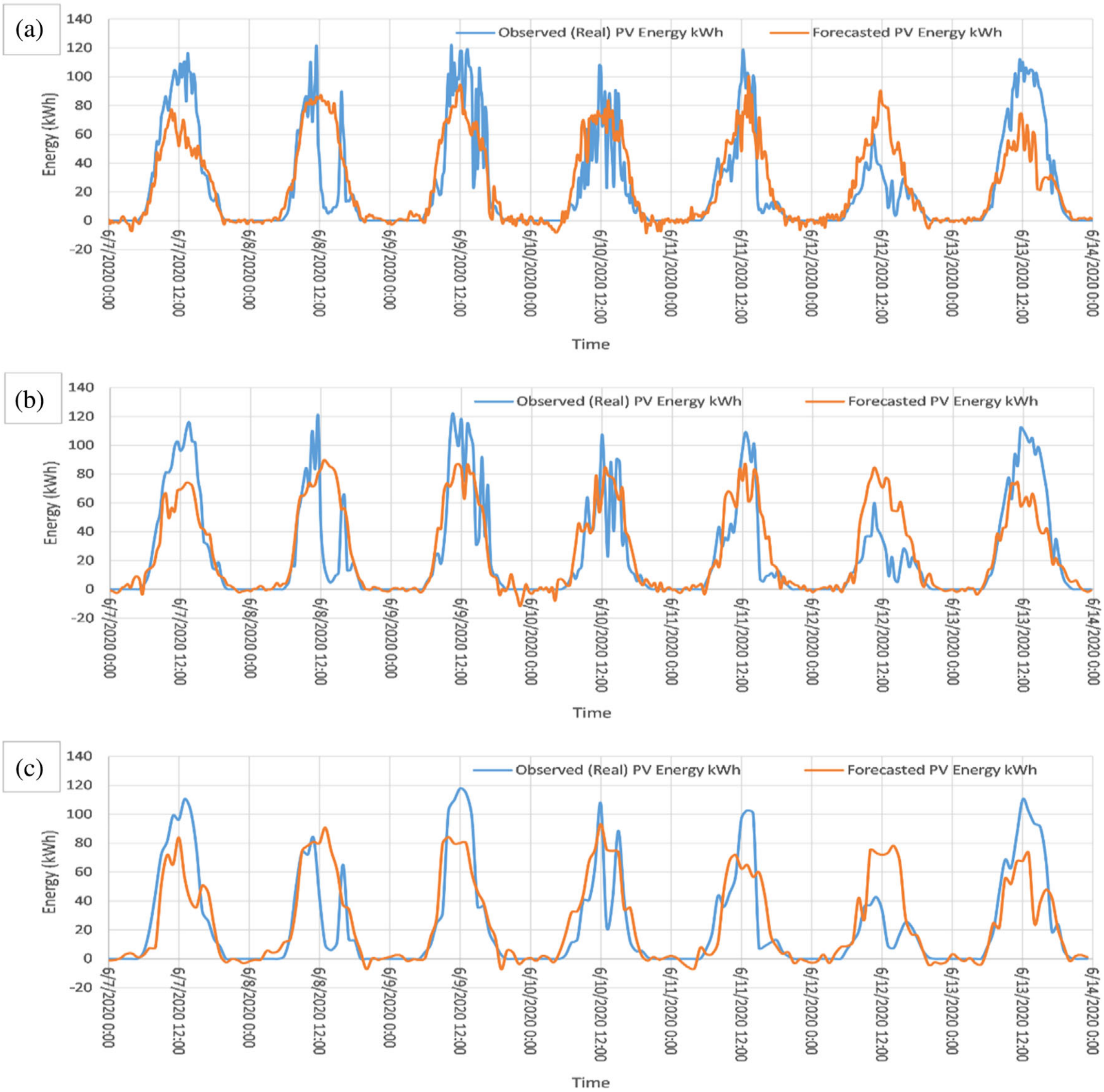

FIGURE 4 PV energy forecasting model performance utilizing (a) 15-; (b) 30-; and (c) 60-min input data resolution

to generate a proper function that solves the forecasting problem, while in the contrast, adding more hidden neurons might result in over-fitting of the training set and, therefore, lowering the ability of generalization.[40] Hence, the number of hidden neurons was set to be 33% (one-third) of the number of inputs. In addition, instead of a time limit, the number of Epochs was set to a limit to control running time. Table 1 shows all settings of ANN parameters depending on input data resolution.

For all ANN models trained here, the data were split into three segments: 70% for the training set, 15% for the validation set, and 15% for the test set. During each epoch, the training set is used to train the models and update the network weights and biases. While

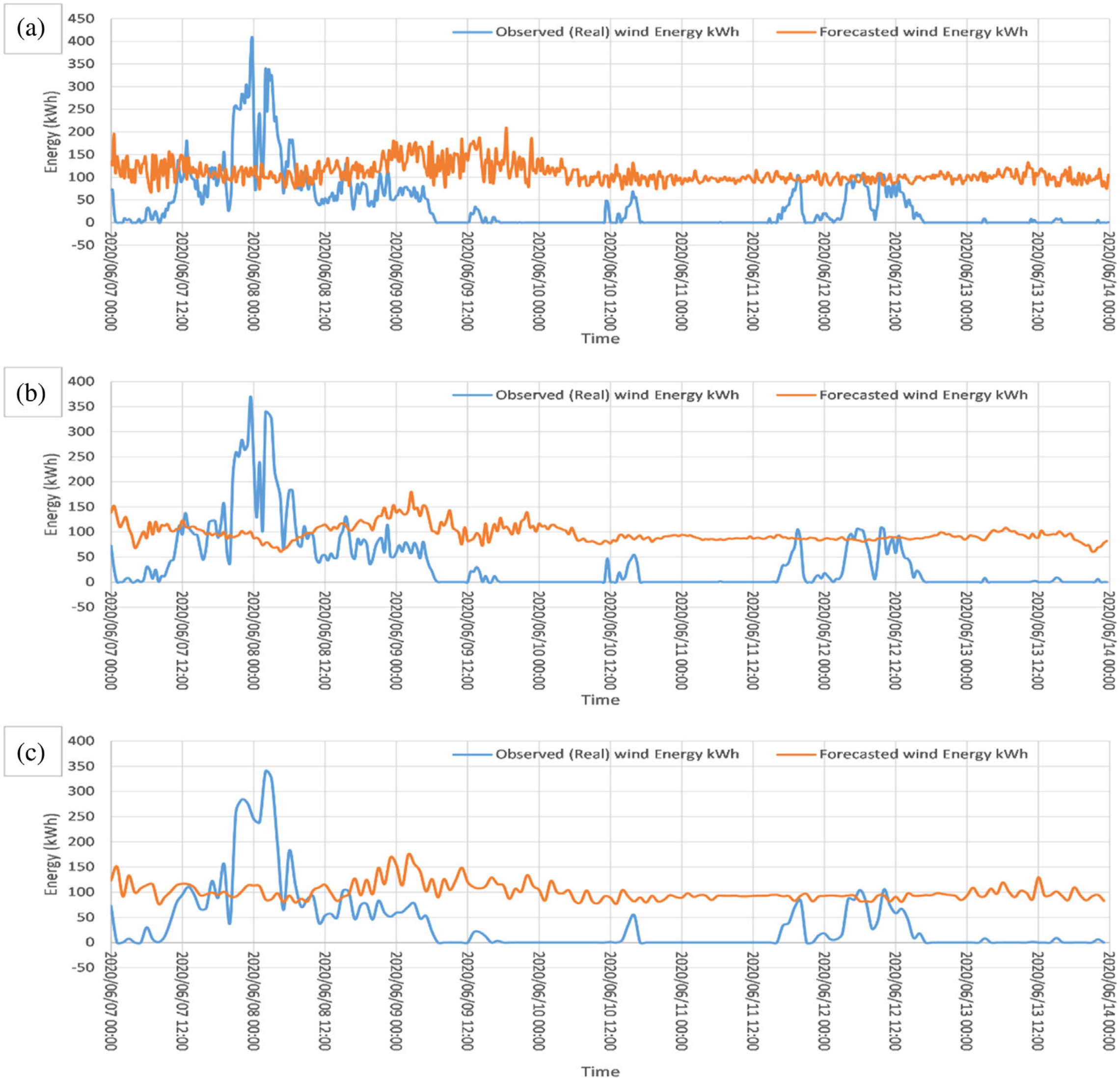

FIGURE 5 Wind energy forecasting model performance utilizing (a) 15-; (b) 30-; and (c) 60-min input data resolution

TABLE 2 Performance measures comparison

Model		Performance measures COD	MAE	MSE
PV forecasting	15-min resolution	0.75	9.72	300.13
	30-min resolution	0.76	10.00	297.81
	60-min resolution	0.74	10.13	312.73
	Average	0.75	9.95	303.55
Wind forecasting	15-min resolution	0.05	119.69	21040.5
	30-min resolution	0.07	116.29	20607.7
	60-min resolution	0.05	116.80	20964.0
	Average	0.06	117.59	20870.7

the validation set is used to monitor the errors during the training process. The training error normally decreases during each epoch, and this applies to the validation set error as well. However, when the network begins to overfit the data, the error on the validation set typically begins to increase. The network weights and biases are saved at the minimum value of the validation set error to ensure that no overfitting has occurred. The test set error is not used during training, but it is used to compare the performance of each epoch. In this work,

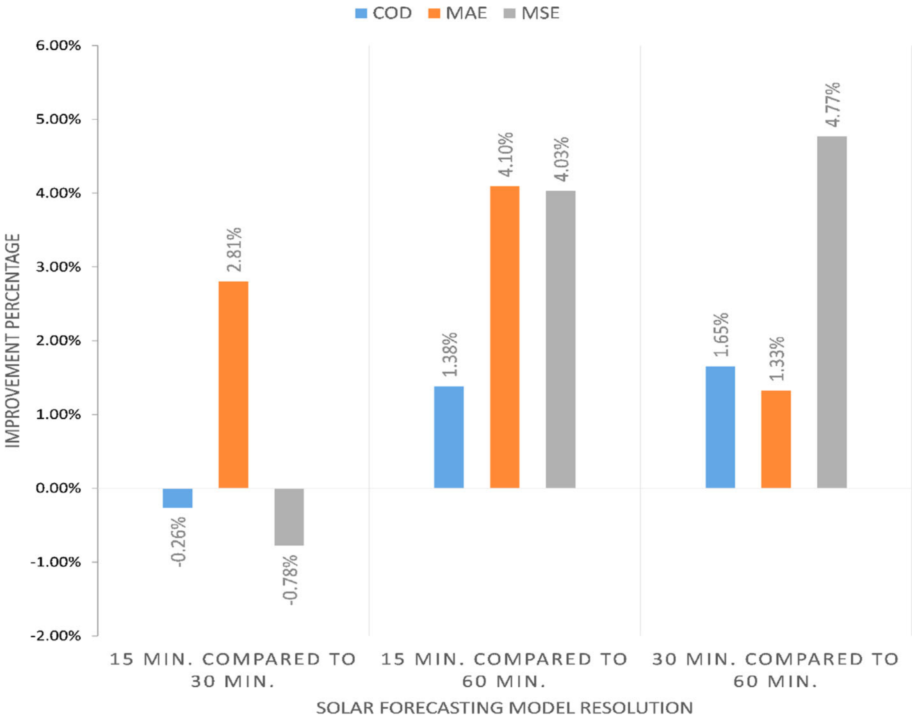

FIGURE 6 Performance measures comparison of PV energy forecasting utilizing different input data resolutions

the performance of each tested model was calculated during the training and testing period, that is, using 85% of the original dataset.

To calculate the performance of the resulting forecasting models, one or more evaluation methods are needed. Evaluation methods such as mean absolute error (MAE), MSE, and coefficient of determination (COD) can be used to evaluate the performance of forecasting models,[20] as can be seen in Equations (12)–(14), where n is the number of observations, y_t is the observed (real) output power at time t, p_t is the forecasted output power at time t, and $\hat{y}$ is the average of the observed values:

$$\mathrm{MAE} = \frac{1}{n} \times \sum_{i=1}^{n} |y_t - p_t| \tag{12}$$

$$\mathrm{MSE} = \frac{1}{n} \times \sum_{i=1}^{n} (y_t - p_t)^2 \tag{13}$$

$$\mathrm{COD} = \frac{\sum_{i=1}^{n} (y_t - p_t)^2}{\sum_{i=1}^{n} (y_t - \hat{y})^2} \tag{14}$$

MAE value is used to measure the closeness between the predicted and measured (real) values. While MSE measures the average squared difference between the predicted values and the measured (real) values. Hence, MSE implies how far the predictions are spread from the measured (real) values. COD (also known as R^2) shows the closeness between the predicted output from the measured (real) data line as a fitted regression. Better modeling accuracy means MAE, and MSE should be closer to zero, while the COD value should be closer to 1. As different measures reflect different views on performance, in this experiment all of the three measures were calculated and compared to evaluate the three PV and three wind models.

3 | RESULTS AND DISCUSSION

The performance of each model was calculated during the training and testing period (one full year). Additionally, as it is difficult to visualize the performance of each tested model for 1 year, the performance of each forecasting model was visualized for the last week of the testing which covers the period June 7–13, 2020 as can be seen in Figure 4 (solar) and Figure 5 (wind).

In Figure 4a, it can be seen that the PV forecasting model utilizing 15 min of input data resolution has good prediction abilities, yet some errors can be observed. Specifically, large errors can be observed on June 8, when the model failed to predict the sudden dip that happened in the afternoon. As the input variables and the forecasted values have high resolution, the small fluctuations in the real energy production can be detected. For instance, days June 9 and 10 show fast fluctuations in the produced energy. The forecasting model was partially able to predict these sudden fast fluctuations, yet it could not accurately predict steep movements.

Figure 4b shows the performance of the PV forecasting model utilizing 30 minutes of input data. This model also shows good prediction abilities. Yet again, some errors can be observed especially on June 8. It can be noticed here that the sudden production fluctuations can still be detected but smaller fluctuations could not be detected as frequently as in the previous model of higher input data resolution. Figure 4c shows the performance of the PV forecasting model utilizing a 60-min resolution of input data. It can be seen that this model has higher forecasting errors for the June 12, 2020. Also, as the input and forecasted data have a lower resolution than the two previous models, production fluctuations appear more smoothly.

Generally, it can be observed from Figure 4 that using ANN forecasting models that only utilize past energy data (time-series past generation data) leads to good forecasting performance. Still, the sudden fluctuations in energy production could not be accurately predicted. This can be explained by the nature of ANN time-series forecasting: models are trained using 1-year past generation data, while the input of the models is the past 24-h data. Based on the trained model and model inputs, each ANN model will forecast the next 24-hour energy generation. As weather fluctuations occur within less than a 24-h window, it is generally hard to accurately predict sudden dips. Forecasting a sudden dip requires either lower forecasting window (lower than the dip, i.e., 1-h forecasting horizon) or real-time dynamic forecasting models where the model can correct the forecast based on the real-time data. Real-time dynamic forecasting has its own applications which are out of the scope of this article.

Moreover, different input data and forecast resolutions show different behavior in detecting and forecasting these fluctuations (as will be discussed later in this section in comparison to wind forecasting).

Figure 5 shows the performance of the wind energy forecasting models utilizing different input data resolutions. Generally, it can be noticed that the ANN time-series forecasting model is not good enough in predicting wind energy.

Figure 5a shows the performance utilizing input data of 15-min resolutions. This wind model was not able to predict the energy accurately, especially in the last few days of testing (June 9–13) when the actual produced energy was zero most of the time. Utilizing input data of 30 and 60 min did not improve the forecasting performance much as can be seen in Figure 5b,c.

In the 15-min resolution forecasting model, the forecasted values were fluctuating in a steeper manner than some of the real-generation values. The steep fluctuations have still existed when utilizing 30-, and 60-min resolutions but in a less-frequent manner.

As the designed ANN-forecasted model only relies on the past-generation time-series data, the output forecasted values were greatly deviating from the real values and huge errors were marked. This indicates that catching seasonality and patterns of wind energy generation by ANN forecasting models requires additional input variables compared to PV energy forecasting.

The results discussed above show a big variance in PV and wind forecasting performance as represented in Table 2. ANN time-series method was efficient in predicting the PV energy output with average COD of 0.75. Also, the average MAE and MSE are 9.95 and 303.55, respectively. The same method with the same data utilization approach shows very poor abilities in forecasting wind energy with a 0.064 COD, 117.59 MAE, and 20870.79 MSE.

Table 2 also shows that ANN time-series method has in general similar abilities in forecasting the PV output energy regardless of the input data resolutions. Although all performance measures are very close and comparable, the 60-min resolution shows higher values of MAE and MSE, yet slightly lower COD. This indicates higher input data resolutions lead to slightly better accuracy—and, interestingly, 30 min performs slightly better than 15 min in some performance measures like COD and MSE. However, the MAE value decreases for higher resolutions indicating slightly better prediction abilities.

With respect to the effect of different input data resolutions on the forecasting model accuracy, it was found that performance measures (as presented in Table 2) are similar to the ones found in the literature. However, this study took an integrated view. For example, the MAE values for day ahead forecasting horizon varies between 7 and 12 depending on the input data and the technique utilized.[41] Similarly, it was also confirmed from the literature that wind

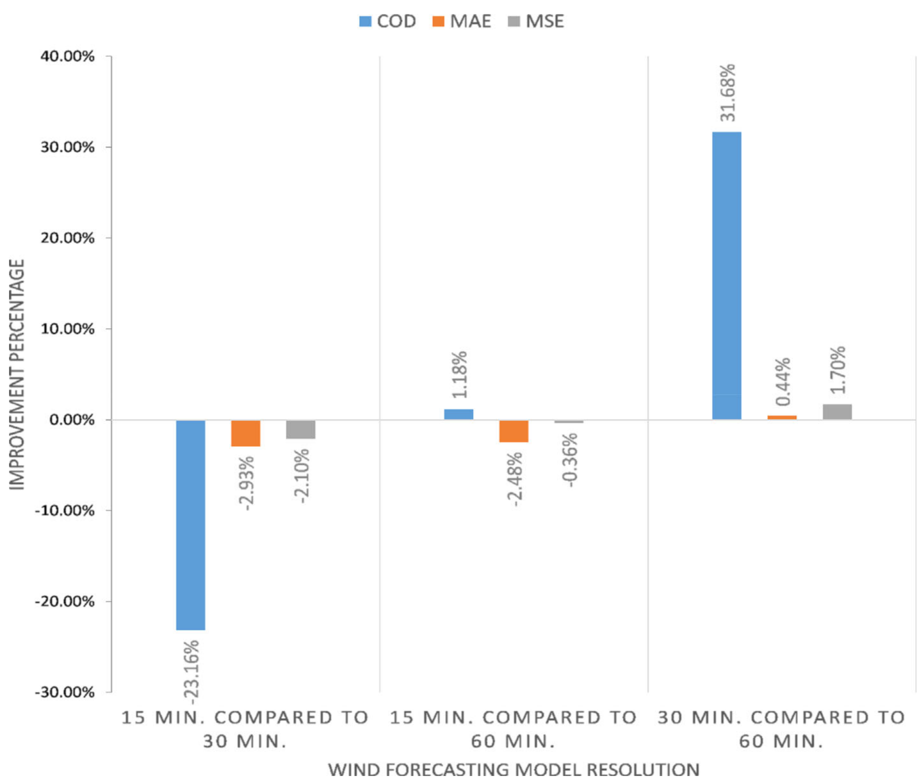

FIGURE 7 Performance measures comparison of wind energy forecasting utilizing different input data resolutions

forecasting models tend to have higher MAE values. Moreover, for some models, the values of MAE do vary greatly between 1 (or even less) up to even a few hundred.[42] Our results are more specific, however, as most other studies only present percentage difference.

COD values are close for all tested resolutions, while some significant differences can be seen between the different utilized resolutions, especially between the 60-min and 15-min resolutions in MAE and MSE. As can be seen from Figure 6, utilizing data with 30-min resolutions instead of 60-min resolutions improved (decreased) MAE by 1.33% and MSE by 4.77%. While utilizing data with 15-min resolutions instead of 60-min resolutions improved MAE by 4.10% and MSE by 4.03%. Utilizing a 15-min resolution instead of 30 does not show any significant improvement. Actually, COD and MSE measures show a deteriorated improvement of 0.26% and 0.78%. Yet MAE shows a 2.81% improvement.

For wind energy forecasting, different input data resolutions show some effects on the forecasting performance. The 30-min resolutions show the lowest MAE and MSE. While higher values of MAE and MSE were observed utilizing 15 min of input data resolution. Interestingly, here 60 min perform better, than 15 min. But every resolution leads to weak performance in general.

As can be seen from Figure 7, utilizing data with 30-min resolutions instead of 60-min resolutions improved COD by 31.68%. Note that even after this huge improvement, COD values for both 30- and 60-min resolutions are still low. Utilizing data with 15-min resolutions instead of 60-min resolutions improved COD by 1.18%, but MAE and MSE did not improve. Utilizing a 15-min resolution instead of 30 does not show any improvement. on the contrary, all the performance measures show deteriorated values.

The results indicate that for intraday renewable energy forecasting, using a 15-min resolution might not lead to the best accuracy for all forecasting purposes. The results also indicate that predicting wind power utilizing the time-series data alongside with ANN method might not lead to good forecasting accuracy at all. As was shown earlier in this section, the ANN shows poor abilities in forecasting wind power utilizing only time-series (past energy) data, while the same method utilizing the same time-series data shows good forecasting accuracy for PV power forecasting. Hence it can be concluded that different renewable energy predictions might require different models, methods, and input data settings. A powerful forecasting method for one renewable energy resource does not necessarily mean that this method is also powerful for forecasting other renewable sources.

4 | CONCLUSION

This article clarified in detail the effects of utilizing different input data resolutions on the performance measures of several forecasting models for both PV and wind farms. The findings are results of building and testing ANN-based PV and wind energy forecasting models using different input data resolutions utilizing real site data. Specifically, although there are claims that higher data resolution leads to better forecasting performance, this article demonstrates that utilizing input data with higher data resolutions might not always lead to better forecasting accuracy compared with lower resolutions.

It was found that ANN time-series model was efficient in predicting the PV energy regardless of the input data resolution. In fact, input data resolutions have only a small effect on the accuracy of the ANN time-series PV forecasting model as forecasting measures are fairly close when utilizing 15 or 30 min input data resolution. Yet, the 15-min resolution shows better forecasting performance compared to the 60-min resolution as it improves some performance measures by 1.3%–4.1%. The same model approach shows poor performance in predicting wind energy. ANN time-series wind forecasting model has huge errors in forecasting wind energy regardless of the input data resolution. Yet, the 30-min input data resolution shows a slightly better performance. Utilizing the 30 min improves some performance measures by 0.4%–31%. These results show that forecasting energy production in a 15-min resolution might not assure high prediction accuracy for all renewable resources. Different renewable energy resources might need different input data resolutions to attain better forecasting accuracy.

AUTHOR CONTRIBUTIONS

Mutaz AlShafeey: Conceptualization (Lead); Data curation (Lead); Formal analysis (Lead); Investigation (Equal); Methodology (Equal); Software (Lead); Validation (Equal); Visualization (Equal); Writing - original draft (Equal); Writing - review & editing (Equal). **Csaba Csaki:** Investigation (equal); methodology (equal); resources (supporting); software (supporting); supervision (lead); validation (equal); visualization (supporting); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

Solar power plant operation data are provided by 3Comm Hungary (3comm.hu/). Wind turbine data is supplied by E.on Hungary (eon.hu/). Project no. TKP2020-NKA-02 has been implemented with support from the National Research, Development and Innovation Fund of Hungary under the "Tématerületi Kiválósági Program" funding scheme.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Due to confidentiality agreements, supporting data can only be made available to researchers upon a non-disclosure agreement.

ORCID

Mutaz AlShafeey https://orcid.org/0000-0002-0935-226X
Csaba Csaki https://orcid.org/0000-0002-8245-1002

REFERENCES

1. Nelson VC, Starcher KL. *Introduction to Renewable Energy*. CRC press; 2015.
2. Perea-Moreno M-A, Hernandez-Escobedo Q, Perea-Moreno A-J. Renewable energy in urban areas: worldwide research trends. *Energies*. 2018;11(3):577.

3. Almutairi A, Sayed K, Albagami N, Abo-Khalil AG, Saleeb H. Multi-port PWM DC-DC power converter for renewable energy applications. *Energies*. 2021;14(12):3490.
4. Alshafeey M, Csáki C. *A case study of grid-connected solar farm control using artificial intelligence genetic algorithm to accommodate peak demand*. IOP Publishing; 2019:012017.
5. Grigsby LL. *Electric Power Generation, Transmission, and Distribution*. CRC press; 2018.
6. Samuelson S, Yang I. Data-driven distributionally robust control of energy storage to manage wind power fluctuations. In: *2017 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE. 2017; 199-204.
7. Karakosta C, Marinakis V, Flamos A, Tuerk A, Frieden D. Expanding RES cooperation with West Balkans: from importing electricity to exporting RES. *Int J Energy Sector Manag*. 2016;10:363-380.
8. The European Parliament and the Council Of The European Union. Regulation (EU) 2019/943 on the Internal Market for Electricity. *Official Journal L*. 2019;158(14.6.2019):54-124.
9. Papapostolou A, Karakosta C, Marinakis V, Flamos A. Assessment of RES cooperation framework between the EU and North Africa. *Int J Energy Sector Manag*. 2016;10:402-426.
10. Boemer J, Burges K, Zolotarev P, et al. Overview of German grid issues and retrofit of photovoltaic power plants in Germany for the prevention of frequency stability problems in abnormal system conditions of the ENTSO-E region Continental Europe. In: *Proceedings of the International Workshop on Integration of Solar Power into Power Systems*. TIB Hannover; 2011:3-8.
11. Wang X, Adelmann P, Reindl T. Use of $LiFePO_4$ batteries in stand-alone solar system. *Energy Procedia*. 2012;25:135-140.
12. Jones L, Zavadil R, Grant W. The future of wind forecasting and utility operations. *IEEE Power Energy Mag*. 2005;3(6):57-64.
13. Lee J, Park G-L, Kim E-H, Kim Y-C, Lee I-W. Wind speed modeling based on artificial neural networks for Jeju area. *Int J Control Autom*. 2012;5(2):81-88.
14. Lerner J, Grundmeyer M, Garvert M. The importance of wind forecasting. *Renew Energy Focus*. 2009;10(2):64-66.
15. Suganthi L, Samuel AA. Energy models for demand forecasting—a review. *Renew Sust Energ Rev*. 2012;16(2):1223-1240.
16. Ahmed A, Khalid M. A review on the selected applications of forecasting models in renewable power systems. *Renew Sust Energ Rev*. 2019;100:9-21.
17. Vagropoulos SI, Chouliaras GI, Kardakos EG, Simoglou CK, Bakirtzis AG. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In: *2016 IEEE International Energy Conference (ENERGYCON)*. IEEE. 2016;1-6.
18. Wang C-H, Grozev G, Seo S. Decomposition and statistical analysis for regional electricity demand forecasting. *Energy*. 2012;41(1):313-325.
19. Lantz B. *Machine Learning with R: Expert Techniques for Predictive Modeling*. Packt Publishing Ltd; 2019.
20. Ahmed R, Sreeram V, Mishra Y, Arif M. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew Sust Energ Rev*. 2020;124:109792.
21. AlShafeey M, Csáki C. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. *Energy Rep*. 2021;7:7601-7614.
22. Aggarwal S, Saini L. Solar energy prediction using linear and non-linear regularization models: a study on AMS (American Meteorological Society) 2013–14 solar energy prediction contest. *Energy*. 2014; 78:247-256.
23. Reikard G. Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Sol Energy*. 2009;83(3):342-349.
24. Chang WY. Wind Energy Conversion System Power Forecasting Using Radial Basis Function Neural Network. *Appl Mech Mater*. 2013; 284:1067-1071.
25. Heinermann J, Kramer O. Machine learning ensembles for wind power prediction. *Renew Energy*. 2016;89:671-679.
26. Liu H, Chen C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl Energy*. 2019; 249:392-408.
27. Demolli H, Dokuz AS, Ecemis A, Gokcek M. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manag*. 2019;198:111823.
28. Liu H, Shi J, Erdem E. An integrated wind power forecasting methodology: interval estimation of wind speed, operation probability of wind turbine, and conditional expected wind power output of a wind farm. *Int J Green Energy*. 2013;10(2):151-176.
29. Wu YK, Hong JS. A literature review of wind forecasting technology in the world. In: *2007 IEEE Lausanne Power Tech Conference Proceedings*. IEEE. 2007;504-509.
30. Hao Y, Tian C. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Appl Energy*. 2019;238:368-383.
31. Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama RJN. A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. *Neurocomputing*. 2019; 365:54-61.
32. Li C, Lin S, Xu F, Liu D, Liu J. Short-term wind power prediction based on data mining technology and improved support vector machine method: A case study in Northwest China. *J Clean Prod*. 2018;205: 909-922.
33. Gallego C, Pinson P, Madsen H, Costa A, Cuerva A. Influence of local wind speed and direction on wind power dynamics–application to offshore very short-term forecasting. *Appl Energy*. 2011;88(11):4087-4096.
34. Wang Y, Hu Q, Meng D, Zhu PJ. Deterministic and probabilistic wind power forecasting using a variational bayesian-based adaptive robust multi-kernel regression model. *Appl Energy*. 2017;208:1097-1112.
35. Khishe M, Mosavi M, Moridi A. Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation. *Appl Acoust*. 2018;137: 121-139.
36. Akkaya B, Çolakoğlu N. Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases. In: *y-BIS 2019 Conference Book: Recent Advances in Data Science and Business Analytics*. Istanbul University; 2019:162-172.
37. Kim P. *Matlab deep learning: With machine learning, neural networks and artificial intelligence*. 1st ed. APress; 2017.
38. Ghritlahre HK, Prasad RK. Application of ANN technique to predict the performance of solar collector systems-a review. *Renew Sust Energ Rev*. 2018;84:75-88.
39. Talaat M, Farahat M, Mansour N, Hatata AJE. Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. *Energy*. 2020;196:117087.
40. Setyawati BR. *Multi-Layer Feed Forward Neural Networks for Foreign Exchange Time Series Forecasting*. West Virginia University; 2005.
41. Rajagukguk RA, Ramadhan RA, Lee HJ. A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. *Energies*. 2020;13(24):6623.
42. Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. *Energy Convers Manag*. 2019;198: 111799.

How to cite this article: AlShafeey M, Csaki C. The impact of input data resolution on neural network forecasting models for wind and photovoltaic energy generation using time series data. *Environ Prog Sustainable Energy*. 2023;42(3):e14039. doi:10.1002/ep.14039