INTRODUCTION

The negative effects of environmental problems such as climate change, pollution, and energy-security have increased the pressure for cleaner and more efficient renewable energy sources. Wind power is a fundamental part of the transition to renewable energy; in 2017, an estimated 17% of all renewable-generated electricity worldwide came from wind sources, and wind energy corresponded to approximately 23% of all the renewable energy production capacity worldwide. The potential for the exploitation of this resource is enormous; it is estimated that by 2050 it will be able to supply approximately 4.4 TW, which corresponds to roughly 37% of all of the end-use power supply in the world.
In Japan, wind energy enjoys less protagonism. Despite having a reported potential of as much as 1900 GW, as of 2017 the installed capacity was only of approximately 3.4 GW. In that same year, the output from wind power sources was roughly 3.7% of all renewable-generated electricity, and merely 0.6% of all generated electricity. In recent years, and particularly after the nuclear power supply was drastically cut due to the Fukushima disaster, the transition to renewable sources has gained momentum and support from both the private and public sectors. However, fossil fuels still represent more than 80% of all energy sources.

The daily output of a turbine depends not only on the cube of the wind speed, but also on the periods when the wind speed is between cut-in and cut-out values. Grid scheduling and electricity trading may also be affected because of the uncertainty and volatility of wind and therefore of the generator output. Moreover, other aspects of operation such as estimation of the life span of turbines and scheduling of maintenance also require reliable wind speed forecasts. That is, optimizing the use of wind energy and integrating it into existing grids relies on accurate knowledge of the future wind conditions, especially in regions with complex topographic features. Such is the case of Japan, whose geography presents additional challenges due to steep, mountainous terrain that modifies the wind profile, introduces turbulence, and causes flow separation and/or flow turning.

In general, wind prediction models can be grouped into statistical and physical models. Physical-based models, or numerical weather prediction (NWP) models, are able to simulate the real state of the atmosphere by using known physical laws. Usually, these models perform better in forecasting upper atmospheric patterns, but errors and biases increase as one gets closer to the surface. On the other hand, pure statistical methods obtain their forecasts from the analysis of recorded historical data and rely mostly on regression techniques. Statistical methods are better at handling local weather patterns and close to the surface patterns, but their performance decreases rapidly as the number of simulated hours increases.

Most implementations of statistical models are in the form of a postprocessing step, to correct and enhance the results of the dynamical NWP model. Kalman filters, analog schemes, Ensemble Model Output Statistics, and quantile regression are among the most common methods for statistical postprocessing of weather forecasts. In recent years, machine learning (ML) and artificial intelligence techniques have also been developed, due to the abundance of forecasted and observed data. Traditional postprocessing methods have been improved by the inclusion of ML components. In the context of ensemble forecasts, Ref. 22 replaced the link functions in a distributional regression model with a neural network able to predict the distribution parameters from a set of predictors, and applied it to 2-m temperature forecasts. The quantile regression method, originally developed by Brempnes, was enhanced by Cannon with the development of a monotone composite quantile regression neural network; and again by Brempnes by using neural networks to estimate the coefficients of Bernstein basis polynomials, after which the quantile function is constructed. On the other hand, Taillardat et al developed instead a random forest generalization for quantile regression, first as a nonparametric approach and posteriorly extended within a semiparametric model. Finally, Chapman et al introduce the use of convolutional neural networks for the correction of gridded, two-dimensional integrated vapor transport forecasts, based on image processing techniques and denoising autoencoders.

For the particular case of wind speed, most ML-related applications have been on prediction based solely on past time series. A common approach would consist of a regression ML model, often in combination with an error correction or optimization component. These implementations lack an NWP component, and therefore, their performance is limited to ultra-short-term applications, rendering them unusable for reliable wind power assessment or wind farm arrangement.

On the other hand, attempts at employing ML for error correction include the use of nonlinear autoregressive exogenous networks, Markov stochastic processes, sequence transfer algorithms, polynomial neural networks, support vector machines, and bidirectional gated recurrent unit neural networks. All of these methods require the output from an NWP model in either deterministic or ensemble form.

Nonetheless, we consider that the literature regarding ML for wind speed error correction is still very limited, and it deserves further exploration. Most literature is not focused on day-ahead wind speed forecasts at hub-height level, which is of crucial importance to the wind farm arrangement. Moreover, just a few of these techniques have been validated in zones with highly fluctuating wind, an important point for the case of Japan where complex topography results in complex wind profiles. Finally, to the best of the authors’ knowledge, no ML-based correction method has considered the inclusion of other weather variables to improve in the correction process, despite being common practice in classical correction methods such as model output statistics (MOS), and despite the versatility of ML algorithms, especially neural networks.

Given these considerations, we present a novel postprocessing correction method for short-term, hub-height (80 m above ground) wind speed predictions from the Weather Research and Forecasting (WRF) model. We propose a neural network architecture of which the inputs are the raw forecasts of multiple variables from the WRF model, and the output is an improved wind speed forecast. The main motivation behind the use of neural networks is their capabilities to recover
nonlinear relationships and dependencies in complex geo-
physical processes, as well as to learn from historical data,
making them ideal for localized prediction and correction.
Similarly to Ref., the proposed method in this paper is also
inspired by the working idea behind denoising autoencoders.
The main contributions of the framework presented in this
paper are summarized in the following points:

• To the best of the authors’ knowledge, this method is the
first to employ a denoising autoencoder-based architec-
ture to correct short-term wind predictions. No complex
decomposition of data is required, and no ensemble results
are needed. The network layout is demand-oriented: scalable and modifiable for time series of different lengths,
or a different number of input variables depending on the
available data.
• It is the first method to combine multiple weather vari-
bles and neural networks to correct wind prediction from
an NWP model to the best of the authors’ knowledge. We
demonstrate that the error correction for a single variable
(wind speed) can be improved significantly by providing
the network with additional information about the atmo-
sphere’s state (wind direction, temperature, and pressure).
We show that the network can find hidden information in
these additional variables by evaluating the final weights
of the trained network.
• The model is validated with data from fifteen sites in an
island of complex geographic features with highly fluc-
tuating winds. Each site is at different elevations and/or
distance from the shore (as shown in Figure 1B), meaning

FIGURE 1 A, Map showing the parent
and nested domains for the WRF model.
The location of the wind farm of interest is
shown with the red triangle in the innermost
domain. B, Elevation [m] map of the wind
farm at Awaji Island showing the location of
the 15 turbines.

(A)

(B)
that each site is influenced by different local wind patterns due to topographic interference. We show that the trained networks are able to learn from these patterns and succeed at reducing the error in all fifteen sites.

The rest of the paper is organized as follows. In Section 2, the information of the physical model and dataset is described. Section 3 shows the details of the proposed multivariable neural network architecture. Section 4 analyzes the performance of our proposed method, and a brief discussion is given. The paper ends with concluding remarks in Section 5.

2 PHYSICAL MODEL AND DATA

This section describes the characteristics of the NWP model used to obtain the forecasts. Throughout this paper, we refer to the direct output from the NWP model as the raw or uncorrected forecast. This section also explains the collection of observed data at hub-height that is used to build and validate the performance of the correction models.

The high-resolution meteorological model chosen in the study is the WRF model with Advanced Research WRF (ARW) version 3.9.1.37 It is a regional model based on a fully compressible and nonhydrostatic dynamic core and can be driven with proper initial/boundary conditions from other large-scale models (e.g., Global Forecast System [GFS] in this study). The forecasting domain was configured by following the recommendations in.38,39 As shown in Figure 1A, one parent (D01) and three nested domains (D02, D03, and D04) with a horizontal resolution of 13.5, 4.5, 1.5, and 0.5 km, respectively, are chosen. There are 35 vertically stretched eta levels for all domains, 10 of which are within the lowest 1 km. The main physical options we used include WRF Single-Moment 6-class (WSM6) microphysics parameterization scheme;40 the new Kain-Fritsch convective parameterization41; the Noah land surface model42; Dudhia shortwave radiation43 and Rapid Radiative Transfer Model longwave radiation44; and the Asymmetric Convective Model version 2 planetary boundary layer scheme.45

GFS forecasting data from the National Center for Environmental Prediction are chosen as the initial and boundary conditions for the regional high-resolution model. The data period is from 2016-01-03 to 2020-06-01 with a 6-hour interval, and a horizontal resolution of 0.5 x 0.5 degrees for all variables. The topographic data were fetched from the US Geological Survey global 30 arc-s elevation (GTOPO30) dataset for all domains.

Predictions are obtained for the same period (2016-01-03 to 2020-06-01). The WRF model is re-initialized as a “cold-start” at 18:00 UTC (03:00 JST) each day and each re-initialization runs for 30 hours. Outputs are produced at 1-hour interval; however, the data for the first 6 hours corresponding to the spin-up time are discarded.37 Hence, the final obtained time series is for the remaining 24 hours between 00:00 UTC (09:00 JST) and 23:00 UTC (08:00 JST).

The nacelle-based wind observations from 15 turbines of the wind farm of interest (red triangle in Figure 1), which is located in the south Awaji area, are used to build and evaluate the performance of the proposed model. Observed data are available every 10 minutes for the same period of obtained GFS data. The data have been quality-controlled to some extent based on the standards addressed in46 to flag out any unreasonable data points. Daily mean wind speed variations and the annual frequency histogram of wind speeds at the wind farm of interest are presented in Figure 2.

3 POSTPROCESSING FORECASTS WITH NEURAL NETWORKS

In this section, we present the architectures for our correction method, the logic behind the design, and the training details. We start with a short explanation of the concept of denoising autoencoders, after which our methods were inspired.
3.1 | Denoising autoencoders

Denoising autoencoders are a particular architecture of neural network, which can remove noise from the input data. The idea consists of a network that maps a vector of noisy data $\tilde{x}$ into a reconstructed vector $\hat{x}$. The network learns the mapping during training by optimizing the parameters of the network to minimize the difference between the clean, noiseless data $x$, and the reconstruction $\hat{x}$ given a cost function $L$. The input and the output layers of a denoising autoencoder are of the same dimensions, and the hidden layers are usually of reduced dimensionality. A sketch of a denoising autoencoder is shown in Figure 3A.

This mapping from corrupted to uncorrupted data can be visualized as learning a manifold.\textsuperscript{47,48} We assume the real information $x$ lies in an abstract mathematical space, or manifold, and that the noisy information $\tilde{x}$ lies outside of this manifold. Hence, the neural network learns the distribution $p(x|\tilde{x})$ and uses it to bring $\tilde{x}$ to the more likely position $\hat{x}$, as shown in Figure 3B.

In this paper, we consider the 24-hour wind speed forecast $\tilde{s} = [\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_{24}]$ from WRF model, and their corresponding observations $s = [s_1, s_2, \ldots, s_{24}]$ to be different by a vector $e$ containing all errors and biases:

$$s = \tilde{s} + e. \quad (1)$$

Therefore, we consider $\tilde{s}$ to be a noisy or corrupted version of $s$ so that by using a network $f$ based on Figure 3A, the corrupted forecast can be mapped to the more accurate estimation $\hat{s}$,

$$\hat{s} = f(\tilde{s}; W), \quad (2)$$

where $W$ are the parameters (weights and biases) of the network. We choose to perform the correction for all 24 values at the same time, so that any temporal relations can be learned by the network. Moreover, if there are any daily periods of high, fluctuating wind where the error might be higher than the daily average, the networks would, in principle, learn to recognize it as well. Finally, since we are training a different network per each site, each network should be able to learn these periods even if they occur at different times of the day in each location.

Although the idea of denoising data with neural networks was originally introduced in\textsuperscript{49} and,\textsuperscript{50} modern denoising autoencoders (presented in\textsuperscript{47} and described in this section) were developed with the distinct intent of finding and extracting meaningful features from data. Naturally, modern denoising autoencoders have been used on actual image and audio denoising tasks such as in\textsuperscript{51} and.\textsuperscript{52}

3.2 | Neural network architecture

We firstly present a single variable network that directly resembles a denoising autoencoder and it is shown in Figure 4A. The input and output layers consist of 24 neurons for $\tilde{s}$ and $\hat{s}$, respectively. The only hidden layer has 23 nodes since all of the performed tests showed that a slight reduction in dimensionality in the hidden layer contributed to enhancing the correction. All layers are fully connected; thus, the output $z_n$ from the $n$th node in the hidden layer is

$$z_n = \phi (\tilde{s}^T w_n + b), \quad (3)$$

where $\phi$ is an activation function, $w_n$ are the weights connecting the input nodes and node $n$, and $b$ is a bias term. The chosen activation functions for the hidden layer and the output layer are ReLU and sigmoid, respectively, as shown in Table 1. We denominate this network as NN-S, where “S” stands for wind speed.

We hypothesize that feeding additional information about the state of the atmosphere into the network will greatly enhance its correction skill. With this idea in mind, we present

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{A. A sketch of a denoising autoencoder. Note that the hidden layers have less nodes than the input and output layers. B. The correction process (Adapted from\textsuperscript{47}). The manifold is presented by the thick black line and the corrupted inputs $\tilde{x}$ by the red dots. The autoencoder displaces each corrupted input $\tilde{x}$ closer to the learned manifold. $x$ and $\tilde{x}$ stand for the noiseless data and a reconstructed vector, respectively.}
\end{figure}
a second network where the input consists of the wind speed forecast \( \tilde{s} \), and also the forecasts for other atmospheric variables, namely wind direction \( \tilde{d} \), pressure \( \tilde{p} \), and temperature \( \tilde{t} \). These extra input variables were chosen on the assumption that they are correlated to the variable of interest. Moreover, this same set of variables is used in other prediction/correction methods where a more comprehensive state of the atmosphere is needed as input. The predictions of these variables are extracted from the same WRF model run as the wind speed prediction, for the same time period.

The architecture of this network is shown in Figure 4B. It has an input layer with 96 nodes for the 24-hour forecast of the four variables, two hidden layers with 24 and 23 nodes, respectively, and an output layer with 24 nodes. The most important feature in this network is that the first hidden layer is not fully connected to the input layer, but rather locally connected. Each node in the first hidden layer is connected only to the input nodes that receive the values for a particular hour. In other words, the \( n \)th node in the hidden layer is connected only to the input nodes for the predicted wind speed \( \tilde{s} \), wind direction \( \tilde{d} \), pressure \( \tilde{p} \), and temperature \( \tilde{t} \) at the \( n \)th hour. Thus, the output \( z_n \) from the \( n \)th node in the hidden layer becomes

\[
z_n = \phi \left( w_{sn} \tilde{s}_n + w_{dn} \tilde{d}_n + w_{pn} \tilde{p}_n + w_{tn} \tilde{t}_n + b \right).
\]

We denominate this network NN-SDPT, where “SDPT” stands for speed, direction, pressure, and temperature. A more detailed description of this network also can be found in Table 1. During training, the network will adjust the connecting weights according on the contribution of each input variable to the correction process. We show in Section 4.2 that the nodes corresponding to the predicted wind will exhibit the largest weights after training, as it is to expect. Each

**FIGURE 4** The two networks used to correct with speed forecasts. A, NN-S: A neural network in which the input is the wind speed forecast. B, NN-SDPT: A neural network in which the inputs are multiple variables including the WRF-predicted wind speed, wind direction, pressure and temperature.
other extra variable contributes in a smaller extent, but their overall contribution results in a significant improvement over network NN-S (as shown in Section 4).

The choice of a locally connected layer instead of a fully connected has the intent of obtaining a regularizing effect, as described in. By choosing to connect only certain nodes between the two layers, we are hard wiring our prior knowledge of the relationships in the data into the network and, at the same time, restricting the weights search space so that the desired weight configuration in the first layer becomes more accessible.

To prove the advantages of a locally connected layer, we have compiled in Table 2 the summarized results of the validation tests carried out to determine the adequate network architecture. We present network NN-SDPT labeled as (A), an equivalent version using a fully connected layer (B), and also a pair of networks where both locally connected and fully connected versions have 24 nodes in the middle layer (C and D, respectively). Each network was trained for all 15 sites using 10-fold cross-validation, and the values in Table 2 represent the average improvement of each metric across all sites and all folds. In all cases, early stopping was used with a patience parameter of 200 epochs, and the network at the epoch showing the lowest validation loss was chosen.

The results show that for both number of nodes, the locally connected versions (A and C) perform better than each of their fully connected counterparts (B and D), with network A being the overall best. To further corroborate the performance of the locally connected version, we carried out a Wilcoxon signed-rank significance test between networks A and B, using all available 150 pairs (15 sites and 10 folds). The test concluded in that there is a significant difference between the values of A and B, with a test statistic $z = -2.60$ and a $P$-value $P < .01$ for root mean square error (RMSE), as well as for CC ($z = 2.04$, $P < .05$) and mean absolute error (MAE; $z = -2.18$, $P < .05$). Note that the results in Table 2 correspond to the average of validation sets, and therefore differ those shown in Section 4.

The remaining hyperparameters such as total number of layers and number of nodes in each layer were determined in an empirical trial-and-error basis until satisfactory performance was obtained.

### 3.3 Training details

For the training of the networks, the dataset was divided such that 90% is used for training and validation, and 10% is used for testing. This corresponds to the periods of 2016-01-03 to 2019-12-24 for training and validation, and 2019-12-25 to 2020-06-01 for testing. The training was done in a 10-fold cross-validation fashion, and the best fold was taken for each site.
We train a different network for each site so that each network can learn local patterns or intricacies in the wind behavior that are important for the correction process. A total of 30 networks are trained, all of them with Adapted Momentum Estimation (ADAM) as optimization and mean square error as the loss function $L$. All inputs are scaled to be between 0 and 1 before feeding them to the networks.

4 | RESULTS

4.1 | Performance of NN-SDPT

In this section, we present results that show the performance for the 15 NN-SDPT networks and also the comparisons with the 15 NN-S networks. We assess the correction skill of the networks by comparing statistical parameters, specifically RMSE, MAE, and correlation coefficient (CC):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(f_i - o_i)^2}{N}},$$  \hspace{1cm} (5)

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N}|f_i - o_i|,$$  \hspace{1cm} (6)

$$\text{CC} = \frac{\sum_{i=1}^{N}(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{(\sum_{i=1}^{N}(f_i - \bar{f})^2)(\sum_{i=1}^{N}(o_i - \bar{o})^2)}},$$  \hspace{1cm} (7)

where $i$ is the time point and $N$ is the total number of verification time points. $f$ and $o$ represent the forecasts (both raw and corrected) and observations, respectively. The bar denotes the mean of the corresponding variable.

As a baseline to assess the neural network's correction skill, we have also included results from the MOS approach in some of the figures. MOS is a family of techniques that corrects the numerical predictions by finding statistical relationships between the NWP's output and observed and/or additional model data. For the MOS implementation shown in this paper, the predictors are the same four weather variables as in NN-SDPT and the observed wind speed values were used as predictands to fit the model. The parameters were determined by using 1 year of past data.

All results shown in this paper correspond to the testing period from 2019-12-25 to 2020-06-01. Values for RMSE, MAE, and CC are shown in Figure 5, Tables 3 and 4, respectively. The “Raw” label corresponds to the parameter calculated between the observations versus the uncorrected WRF wind speed forecasts, while “NN-S” and “NN-SDPT” correspond to the observations versus the forecasts corrected with each method, respectively. The percentage of improvement relative to the raw forecast is also shown for each parameter.

Figure 5 shows the RMSE values before and after correction, as well as the percentage of improvement. For all 15 sites, both networks are able to reduce the RMSE in the testing period. This behavior was to be expected since the loss of function that the network was trained to minimize was the mean square error. The parameters MAE and CC are also improved in all 15 sites. Furthermore, all metrics regarding the correction with the multivariable network, that is, NN-SDPT, are superior to those of the single variable NN-S. This indicates that NN-SDPT is indeed capable to extract additional information from wind direction, pressure, and temperature forecasts that help in correcting the wind speed variable.

Specifically, for the correction with NN-S, the percentage of improvement of the RMSE metric ranges from 6.86% in site 5 to 17.63% in site 1, with an average of 11.55%. By using NN-SDPT, the relative improvement increases with values ranging from 8.76% in site 4 to 26.55% in site 15, with an average of 16.52%. Despite NN-SDPT performing better than NN-S in all sites, the sites where NN-S performs the worst (sites 4, 5, and 13) are also the sites where NN-SDPT performs the worst. A closer look at Figure 5 and Tables 3 and 4 reveals that the raw predictions for these sites are much better than other sites (e.g., site 5 has the lowest raw RMSE and MAE, and the highest CC), so it can be implied that in these sites there is little room for correction when compared to the other sites.

Figure 6 (lower panel) shows the RMSE values calculated for each of the 24 simulated hours, for all sites. The metric for an additional correction with MOS is included for reference (orange dashed line). The RMSE for the raw predictions, displayed with the solid red line, shows small values during the first hours of the simulation, and it grows steadily until its peak at the 12th simulation hour (20:00 JST). The boxplots in the upper panel correspond to the observations at each respective hours, and they show that there is a period of change to strong winds which extends from the 9th hour (17:00 JST).
to the 12th (20:00 JST) and may be the reason of the increased error. A per-site analysis revealed that this high wind speed period exists in all sites, although it is more marked in some sites than in others. After this period, the error decreases slightly before increasing again during the last hours of the simulation. The reduced RMSE after the correction with NN-S and NN-SDPT is shown with the green and blue lines, respectively. The error is largely reduced for all hours and in similar proportions, resulting in a very similar pattern.

### TABLE 3
The MAE values [m/s] for all 15 sites, before (“Raw”) and after (i.e., “NN-S” and “NN-SDPT”) correction

| Site | Raw  | NN-S | % Improv. | NN-SDPT | % Improv. | NN-SDPT |
|------|------|------|-----------|---------|-----------|---------|
| No.1 | 2.40 | 2.00 | 16.47     | 1.87    | 21.85     |         |
| No.2 | 2.27 | 2.00 | 11.92     | 1.94    | 14.48     |         |
| No.3 | 2.04 | 1.86 | 8.64      | 1.79    | 11.94     |         |
| No.4 | 2.10 | 1.97 | 10.02     | 1.94    | 7.56      |         |
| No.5 | 1.92 | 1.81 | 5.55      | 1.75    | 8.77      |         |
| No.6 | 2.22 | 1.94 | 12.56     | 1.73    | 22.09     |         |
| No.7 | 2.18 | 2.06 | 5.52      | 1.90    | 12.89     |         |
| No.8 | 2.09 | 1.83 | 12.50     | 1.76    | 15.78     |         |
| No.9 | 2.15 | 1.74 | 19.34     | 1.65    | 23.33     |         |
| No.10| 2.15 | 1.80 | 16.24     | 1.56    | 27.72     |         |
| No.11| 2.06 | 1.73 | 16.22     | 1.68    | 18.38     |         |
| No.12| 2.17 | 1.99 | 8.09      | 1.76    | 18.56     |         |
| No.13| 2.10 | 1.96 | 6.70      | 1.89    | 9.76      |         |
| No.14| 2.10 | 1.92 | 8.64      | 1.76    | 16.27     |         |
| No.15| 2.24 | 2.09 | 6.64      | 1.89    | 15.37     |         |

### TABLE 4
Same as Table 3, but for CC

| Site  | Raw   | NN-S | % Improv. | NN-SDPT | % Improv. | NN-SDPT |
|-------|-------|------|-----------|---------|-----------|---------|
| No.1  | 0.624 | 0.741| 18.88     | 0.778   | 25.31     |         |
| No.2  | 0.601 | 0.695| 15.73     | 0.719   | 19.79     |         |
| No.3  | 0.695 | 0.732| 5.29      | 0.753   | 8.19      |         |
| No.4  | 0.706 | 0.747| 5.82      | 0.753   | 6.92      |         |
| No.5  | 0.742 | 0.780| 5.22      | 0.790   | 6.50      |         |
| No.6  | 0.686 | 0.740| 8.58      | 0.790   | 16.03     |         |
| No.7  | 0.706 | 0.755| 6.94      | 0.774   | 9.81      |         |
| No.8  | 0.694 | 0.747| 7.59      | 0.767   | 10.48     |         |
| No.9  | 0.680 | 0.742| 8.93      | 0.782   | 14.66     |         |
| No.10 | 0.670 | 0.718| 7.40      | 0.794   | 18.75     |         |
| No.11 | 0.695 | 0.749| 7.74      | 0.770   | 10.72     |         |
| No.12 | 0.676 | 0.723| 6.83      | 0.783   | 15.70     |         |
| No.13 | 0.705 | 0.754| 7.04      | 0.767   | 8.88      |         |
| No.14 | 0.657 | 0.680| 3.50      | 0.744   | 13.16     |         |
| No.15 | 0.685 | 0.737| 7.73      | 0.774   | 13.15     |         |
shown in Figure 6, NN-SDPT outperforms NN-S at all times, and this difference in performance is more evident during the last hours of the simulation.

The Taylor diagrams displayed in Figures 7 and 8 also show the improvement in the forecasts after postprocessing with the neural networks. Figure 7 displays the overall improvement at all 15 sites and includes the MOS correction as well for reference. Figure 8B shows the displacement of the points for each site individually. As a general pattern, correcting with NN-S reduces the centered RMSE (CRMSE) and increases the CC slightly. However, points tend to move toward zero in the standard deviation axis. This means that a great amount of variability in the data is lost during the correction. On the other hand, correcting with NN-SDPT appears to displace the points in a direction more toward the reference points (not shown in 8a). This results in an even higher reduction in CRMSE and a higher increase in the CC. Nevertheless, after correction with NN-SDPT there is still a slight reduction in standard deviation.

Finally, in Figure 9 we present five cases of corrected and uncorrected wind speed time series, together with MOS correction and observed values. Each case lasts a period of 7 days. The metrics for each case and correction method are shown in Table 5. These cases were selected in order to highlight the advantages of the current method when compared to the raw output and MOS, and also to find the any weaknesses of the method that can be improved.

A and B are cases where the raw forecast incorrectly predicts brief periods of high-speed wind. These periods last approximately 1 day. During these periods, MOS offers moderate improvement and sits mostly between the raw and proposed methods, although in case B, MOS can get closer to NN-S. On the other hand, NN-SDPT successfully corrects the overestimation in A and is able to bring the values closer to the observations in case B, surpassing both MOS and NN-S. Case A also shows a period of sustained high-speed wind lasting the last 2 days of the sample. The WRF model underestimated this period, and out of all the tested methods, it could only be corrected with NN-SDPT, with MOS degrading the results for this period. The RMSE values in Table 5 show that for both cases A and B, all three correction methods offer significant improvement with NN-SDPT being superior to MOS and NN-S.

Case C presents another high-speed wind period lasting approximately 3 days. It is interesting to see that MOS and NN-SDPT behave similarly; NN-SDPT performs better during the first day, but MOS is clearly superior in the next 2 days. Cases D and E show examples where the uncorrected values are already very close to the observations. In these instances, a correction method could easily move the prediction farther away from the real values and increase the RMSE, as
it is the case for both MOS and NN-S in case D (Table 5). In that same case, NN-SDPT is able to achieve a slight reduction of RMSE of approximately 7%. On the other hand, the values for case E show the same improvement of 14.22% for both MOS and NN-SDPT, despite the time series for both looking considerably different, with MOS showing higher variability.

This shortcoming of the correction method should be acknowledged. Network NN-SDPT indeed reduces the
RMSE in all of the selected cases, but the corrected values present less variability and randomness, which is a very important characteristic in the behavior of wind. Quick fluctuations in wind speed that might be correctly predicted by the NWP model might be flattened out by the correction process.

**FIGURE 9** Wind speed observations and forecasts (corrected and uncorrected) for a period of 7 days, for five selected cases. A–E represent the Case A–E displayed in Table 5.
Traditionally, it has been common to treat neural network models as black boxes and to give priority only to the accuracy of its output. However, analyzing the inner workings of a network can help determine which features contribute more to the network’s output. This kind of analysis can also provide new insights into complex, nonlinear physical systems. In this section, we attempt to study the internal state of the network NN-SDPT, particularly the connections between the input and the first hidden layer. As shown in Equation 4, each node in the first hidden layer contains a weight for each atmospheric variable. The higher the magnitude of the weight, the higher the respective variable contributes to the correction process.

In order to study the contribution of each variable, we have extracted the final weights from the links between the input and first hidden layer. These are shown in Figure 10 after normalization and averaging for each site. As expected, wind speed has the highest weight in the correction process for all sites except for sites 2 and 4. This means that wind speed is the variable that the network finds the most useful for the correction of wind speed. Although this might seem trivial and obvious from a human perspective, it should be mentioned that the network has no previous knowledge of which variable is being input in each node, and it knows no physical information about the atmospheric processes. The fact that the network precisely figured out the variable that was being corrected suggests that the correction is not happening by pure chance.

In almost all sites, wind speed is followed by temperature, pressure, and, finally, wind direction. It should be noted that this does not necessarily mean that there is a higher correlation between wind speed and temperature than with the other variables. It simply means that the model has found useful information in the temperature time series that improves the correction of the wind speed time series. Although there might be deeper connections and relations in the subsequent layers that we are not taking into account, this brief analysis shows that the correction in each site depends on each variable with different magnitudes, suggesting that each site is certainly subject to very different wind patterns. However, it should be mentioned that any attempt at understanding the weights of a neural network can only be done speculatively, as there can be different interpretations of their meanings.

5 | CONCLUSIONS

This work focused on the challenges presented when forecasting wind speed in regions of complex geographical features where the patterns and behavior of wind are more fluctuating than usual. Intending to improve wind power output prediction in a wind farm in Awaji Island, Japan, we have proposed a novel ML-based correction method, that is, NN-SDPT, to improve the forecasts of wind speed out of the WRF model. The correction process executed by the network is based on extracting underlying information from
the predictions of additional atmospheric variables (wind direction, temperature, and pressure), forecasted by the same WRF model.

It was observed that NN-SDPT has remarkable capabilities to improve the raw forecasts of the WRF model. The 15-site averaged improvements of RMSE, MAE, and CC are 16.52%, 16.32%, and 13.20%, respectively. Moreover, the multivariable network NN-SDPT was shown to be superior to its single-variable counterpart (NN-S) in all tests and metrics, a fact that highlights the existence of useful, hidden information in other atmospheric variables, and the ability of the neural network to find this information.

Although we have presented these networks in the context of wind speed forecasting and wind power generation, we can say without a doubt that these can be used for the correction of other atmospheric variables. The layout of NN-SDPT is easily adaptable and can fit in any weather prediction pipeline, as a correction step for results coming from an NWP model. Finally, since each network is trained locally, this method can be implemented wherever observed and predicted data are available.

ACKNOWLEDGEMENTS

We acknowledge the support of Clean Energy Factory Co., Ltd., Tokyo, Japan, for providing the observation data. This work was partly supported by National Key Research and Development Program of China (grants no. 2018YFC1507005), National Natural Science Foundation of China (grants no. 41905084), and Scientific Research Fund of Chengdu University of Information Technology (grants no. KYTZ201728).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Yuzhang Che https://orcid.org/0000-0001-6335-9326

REFERENCES

1. Jacobson MZ, Delucchi MA, Bauer ZAF, et al. 100% Clean and renewable wind, water, and sunlight all-sector energy roadmaps for 139 countries of the world. Joule. 2017;1(1):108-121.
2. Ohlenforst K, Sawyer S, Dutton A, et al. Global Wind Report 2018. Brussels: Global Wind Energy Council; 2019.
3. IEA. Electricity Information. Paris: International Energy Agency. 2019.
4. IRENA. Renewable Energy Statistics 2020. Abu Dhabi: The International Renewable Energy Agency; 2020.
5. Chen WM, Kim H, Yamaguchi H. Renewable energy in eastern Asia: renewable energy policy review and comparative SWOT analysis for promoting renewable energy in Japan, South Korea, and Taiwan. Energy Pol. 2014;74:319-329.
6. Betz A. Introduction to the Theory of Flow Machines. Pergamon; 1966.
7. Jiang X, Chen H, Xiang T. Assessing the effect of wind power peaking characteristics on the maximum penetration level of wind power. IET Gener Transm Distrib. 2015;9(16):2466-2473.
8. Ding M, Zhou H, Xie H, Wu M, Nakashima Y, Yokoyama R. A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing. 2019;365:54-61.
9. Dunstan T, Lock A, Skea A. Empirical error correction and feature identification for long term wind resource assessment using support vector regression. J Renew Sustain Energy. 2016;8(1):013305.
10. Han X, Liu D, Xu C, Shen WZ. Atmospheric stability and topography effects on wind turbine performance and wake properties in complex terrain. Renew Energy. 2018;126:640-651.
11. Politis ES, Prospathopoulos J, Cabezon D, Hansen KS, Chaviaropoulos PK, Barthelmie RJ. Modeling wake effects in large wind farms in complex terrain: the problem, the methods and the issues. Wind Energy. 2012;15(1):161-182.
12. Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. Renew Sustain Energy Rev. 2008;12:1725-1744.
13. Vannitsem S. Dynamical properties of MOS forecasts: analysis of the ECMWF operational forecasting system. Weather Forecast. 2008;23:1032-1043.
14. Wilks DS. Statistical Methods in the Atmospheric Sciences, 3rd edn. Cambridge: Academic Press; 2011.
15. Zjavka L. Wind speed forecast correction models using polynomial neural networks. Renew Energy. 2015;83:998-1006.
16. Liu H, Tian HQ, Chen C, YFLL. A hybrid statistical method to predict wind speed and wind power. Renew Energy. 2010;35:1857-1861.
17. Kalman RE. A new approach to linear filtering and prediction problems. J Fluids Eng Trans ASME. 1960;82:35-45.
18. Monache LD, Nipen T, Liu Y, Roux G, Stull R. Kalman filter and analog schemes to postprocess numerical weather predictions. Mon Weather Rev. 2011;139:3554-3570.
19. Monache LD, Anthony Eckel F, Rife DL, Nagarajan B, Searight K. Probabilistic weather prediction with an analog ensemble. Mon Weather Rev. 2013;141:3498-3516.
20. Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon Weather Rev. 2005;133:1098-1118.
21. Bremnes JB. Probabilistic forecasts of precipitation in terms of quantiles using NWPModel output. Mon Weather Rev. 2004;132:338-347.
22. Rasp S, Lerch S. Neural networks for postprocessing ensemble weather forecasts. Mon Weather Rev. 2018;146:3885-3900.
23. Cannon AJ. Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. Stoch Environ Res Risk Assess. 2018;32:3207-3225.
24. Bremnes JB. Ensemble postprocessing using quantile function regression based on neural networks and bernstein polynomials. Mon Weather Rev. 2020;148:403-414.
25. Taillardat M, Mestre O, Zamo M, Naveau P. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. Mon Weather Rev. 2016;144:2375-2393.
26. Taillardat M, Mestre O. From research to applications—examples of operational ensemble post-processing in France using machine learning. Nonlinear Process Geophys. 2020;27(2):329-347.
27. Chapman WE, Subramanian AC, Delle Monache L, Xie SP, Ralph FM. Improving atmospheric river forecasts with machine learning. Geophys Res Lett. 2019;46:10627-10635.

28. Demolli H, Dokuz AS, Ecemis A, Gokcek M. Wind power forecasting based on daily-wind speed data using machine learning algorithms. Energy Convers Manage. 2019;198:111823.

29. Jaseena KU, Kovoor BC. A hybrid wind speed forecasting model using stacked autoencoder and LSTM. J Renew Sustain Energy. 2020;12(2):023302.

30. Yu C, Li Y, Bao Y, Tang H, Zhai G. A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manage. 2018;178:137-145.

31. Li R, Jin Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. Appl Energy. 2018;228:2207-2220.

32. Wang L, Li X, Bai Y. Short-term wind speed prediction using an extreme learning machine model with error correction. Energy Convers Manage. 2018;162:239-250.

33. Zhao X, Liu J, Yu D, Chang J. One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data. Energy Convers Manage. 2018;164:560-569.

34. Zhao J, Wang J, Guo Z, Guo Y, Lin W, Lin Y. Multi-step wind speed forecasting based on numerical simulations and an optimized stochastic ensemble method. Appl Energy. 2019;255:113833.

35. Wang H, Han S, Liu Y, Yan J, Li L. Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system. Appl Energy. 2019;237:1-10.

36. Makarynsky O, Makarynska D, Rusu E, Gavrilov A. Filling gaps in wave records with artificial neural networks. In Proceedings of the 12th International Congress of the International Maritime Association of the Mediterranean, IMAM 2005—Maritime Transportation and Exploitation of Ocean and Coastal Resources; 2005:1085-1091.

37. Wang W, Bruyère C, Duda M, et al. ARW Users Guide V3.9-3. 2017.

38. Warner TT, Peterson RA, Treadon RE. A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. Bull Am Meteor Soc. 1997;78(11):2599-2617.

39. Warner TT. Quality assurance in atmospheric modeling. Bull Am Meteor Soc. 2011;92(12):1601-1610.

40. Hong SY, Noh Y, Dudhia J. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon Weather Rev. 2006;134:2318-2341.

41. Kain JS. The Kain-Fritsch convective parameterization: an update. J Appl Meteorol. 2004;43:170-181.

42. Chen F, Dudhia J. Coupling an advanced land surface – HydrologyModel with the Penn State – NCARMM5Modeling System. Part I: model implementation and sensitivity. Mon Weather Rev. 2001;129:569-585.

43. Dudhia J. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J Atmos Sci. 1989;46(20):3077-3107.

44. Mlawer EJ, Taubman J, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res. 1997;102:16663-16682.

45. Pleim JE. A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: model description and testing. J Appl Meteorol Climatol. 2007;46:1383-1395.

46. Freedman J, Zack J, Schroeder J, et al. The Wind Forecast Improvement Project (WFIP): A Public/Private Partnership for Improving Short Term Wind Energy Forecasts and Quantifying the Benefits of Utility Operations – The Southern Study Area. U.S. Department of Energy; 2014.

47. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. Proceedings of the 25th international conference on Machine learning. 2008:1096-1103.

48. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016. http://www.deeplearningbook.org

49. Le Cun Y, Fogelman-Soulié F. Modèles connexionnistes de l’apprétissage. Intellectica Revue de l’Association pour la Recherche Cognitive. 1987;2:114-143.

50. Gallinari P, Le Cun Y, Thiria S, Fogelman-Soulié F. Mémoires associatives distribuées: une comparaison; 1987.

51. Gondara L. Medical image denoising using convolutional denoising autoencoders. In: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW); 2016:241-246.

52. Grais EM, Plumbley MD. Single channel audio source separation using convolutional denoising autoencoders. In: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP); 2017:1265-1269.

53. Kukacka J, Golkov V, Cremers D. Regularization for deep learning: a taxonomy. ArXiv. 2017;abs/1710.10686.

54. Kingma DP, Ba JL. ADAM: a method for stochastic optimization. ArXiv. 2015.

55. Glaun HR, Lowry DA. The use of model output statistics (MOS) in objective weather forecasting. J Appl Meteorol. 1972;11:1203-1211.

56. Carter GM, Dallavalle JP, Glahn HR. Statistical forecast based on the national meteorological center’s numerical weather prediction system. Weather Forecast. 1989;4:401-412.

57. Montavon G, Samek W, Müller KR. Methods for interpreting and visualizing deep neural networks. Digital Signal Process. 2018;73:1-15.

How to cite this article: Salazar AA, Che Y, Zheng J, Xiao F. Multivariable neural network to postprocess short-term, hub-height wind forecasts. Energy Sci Eng. 2022;10:2561–2575. doi:10.1002/ese.928