Defence R&D Canada
External Literature
DRDC-RDDC-2015-P094
October 2015

Short-Term Wind Speed and Power Forecasting using an Ensemble of Mixture Density Neural Networks

Zhongxian Men[a,c,*], Eugene Yee[b,c], Fue-Sang Lien[a,c], Deyong Wen[a,c], Yongsheng Chen[d]

[a] *Waterloo CFD Engineering Consulting Inc., Waterloo, Ontario, Canada N2T 2N7*
[b] *Defence Research and Development Canada, Suffield Research Centre, P.O. Box 4000 Stn Main, Medicine Hat, Alberta, Canada T1A 8K6*
[c] *Department of Mechanical & Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada*
[d] *Department of Earth & Space Science & Engineering, York University, North York, Ontario, Canada*

Abstract

An ensemble of mixture density neural networks is used for short-term wind speed and power forecasting. Predicted wind speeds obtained from a numerical weather prediction model are used as the input data for the mixture density network, whose outputs are the mixture density parameters (used to represent the probability density function of the uncertain output or target variable). All mixture density neural networks in an ensemble are assumed to have a three-layer architecture, with each architecture having different numbers of nodes in the hidden layer. Because a mixture of Gaussian distributions is used to approximate the conditional distribution of the target random variable (either wind speed or wind turbine power), the uncertainties arising from both the model structure and model output can be completely quantified. In consequence, rigorous confidence intervals reflecting these sources of uncertainty in the prediction can be obtained and used to assess the performance for the wind speed and wind turbine power forecasting. An application of the proposed approach to a data set of the measured wind speed and power from an operational wind turbine in a wind farm in Taiwan is used to test the methodology. The results of this application demonstrate that the proposed methodology works well for the multi-step ahead wind speed and power forecasting.

Keywords: Ensemble forecasting, Gaussian mixture, Mixture density neural network, Wind speed/power prediction

1. Introduction

Over the past decade, the generation of clean energy from wind turbines has become ever more prevalent. To support this effort, researchers have focused on improving the capability to forecast the

*Corresponding author
Email address: zmen@uwaterloo.ca (Zhongxian Men)

wind resources (e.g., wind speed and the concomitant wind turbine power generation). Numerous methodologies have been proposed for wind speed and power forecasting, including statistical modeling, physics-based weather forecast modeling, and machine learning approaches. Machine learning techniques, which include artificial neural networks (ANNs) and support vector machines (SVMs), have been extensively applied for wind energy forecasting. Very recently, methodologies used in financial time series modeling such as the autoregressive conditional heteroskedasticity (ARCH) model of Engle [1] and the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev [2] have been applied to wind energy prediction. Traditionally, short-term forecasts of the wind speed and wind power have been based on point forecasts [3,4], but the trend recently has been focussed on the development of probabilistic forecasting of wind resources [5–10]. In particular, Gneiting et al. [10] proposed the use of a calibrated probabilistic forecasting methodology based on a regime-switching space-time model that can account explicitly for alternating atmospheric regimes. For general reviews of the various approaches that have been used for wind speed and wind turbine power prediction, the reader can consult Giebel et al. [11] and Foley et al. [12] and the references therein.

In the applications of these approaches to wind speed and power forecasting, a major theme was the uncertainty quantification for the forecasts. In this regard, ensemble methods have been proposed for wind speed and power prediction (e.g., Men et al. [8,9], Wang et al. [13], among others). One of the advantages in using an ensemble approach is that the confidence intervals for the forecasted wind speeds and powers can be obtained rigorously and then used as criteria for the forecast assessment. Alternatively, Bremnes [6] used local quantile regression for probabilistic forecasting. Gneiting et al. [10] used a truncated Gaussian predictive distribution for probabilistic forecasting of the hourly-averaged wind speeds at a single site. In the latter study, a proper scoring rule referred to as the continuous ranked probability score was utilized to assess the quality of the wind speed forecasts. In this paper, we propose to use an ensemble of mixture density neural networks for the probabilistic forecasting of the wind speed and power. In this approach, the probability density function of the target variable (wind speed or wind turbine power) is represented by a mixture of known probability distributions (herein taken to be Gaussian distributions).

Mixture distributions have been widely utilized in statistics to approximate complex (non-analytical) distributions. It is well known that a mixture of Gaussian distributions can approximate any continuous probability distribution with any accuracy provided the number of components (ker-

nels) in the mixture is sufficiently large. In the seminal work of Bishop [14] and elaborated in greater detail later in Bishop [15], the concept of a mixture density neural network (MDN) was introduced. Here, the probability density function of a target vector of the neural network conditioned on an input vector is represented by a mixture (sum) of Gaussian distributions. The conditional mixture distribution embodies the uncertainty associated with the target vector and allows one to quantify rigorously the uncertainty in the prediction of the target vector.

In this paper, we generalize the approach proposed by Bishop [14,15] to include a quantification of the uncertainty arising from the neural network structure used to approximate the values for the parameters of the MDN, as well as the number of kernels required to adequately represent the probability density function of the target variable. To this end, we formulate an ensemble of MDNs and use the information embodied in this ensemble to obtain the best prediction for the target vector as well as the uncertainty in this prediction. The proposed extension of MDN theory provides a semi-automatic procedure for "selecting" the appropriate number of kernels (required to represent the distribution of the target variable) and hidden units in the MDN (required to represent the mapping from the deterministic input variable to the probability density function of the uncertain target variable).

The paper is organized as follows. Section 2 briefly introduces the concept of a neural network and then describes the approximation of a target vector using a mixture of Gaussian distributions. An ensemble of MDNs is proposed for wind speed and wind turbine power forecasting. Section 3 discusses how to assess the prediction performance by using some well known assessment measures. Section 4 provides an application of the proposed methodology to a data set of wind speed and wind turbine power obtained from a wind farm in Taiwan. This is followed by a brief summary of the results and a conclusion in Section 5.

2. Ensemble mixture density network

It is well known that modeling the relationship between two random variables (or, vectors) is a difficult problem in general. The relationship between any two random variables could be a linear function, a polynomial function, or even a much more complex mapping. In wind turbine power forecasting applications, this unknown mapping (between the predicted wind speed and wind turbine power) is usually represented non-parametrically using a feedforward multilayer neural network. Flexibility is achieved through the use of an essentially non-parametric representation for the map-

ping and wind turbine power prediction can be carried out based on the trained neural network. Before introducing our ensemble MDN approach, we present a very brief review of neural network modeling.

Let x and y represent an independent and dependent random variable (vector), respectively. A general mapping from x to y can be described by

$$y = g(x|\theta) + e, \tag{1}$$

where $g(x|\theta)$, parametrized by a parameter vector θ, is a general model function mapping x to y and e is the residual characterizing the accuracy of using $g(x|\theta)$ to approximate y. In this paper, y is either the measured wind speed at the hub height of a wind turbine or the corresponding wind power generated by the turbine with a specific realization (sample) given by $\mathbf{y} = \{y_1, y_2, \ldots, y_T\}$, where T is the sample size. The independent (input) vector x is a sample of the modeled (predicted) wind speed obtained from a numerical weather prediction (NWP) model. A specific realization (sample) of x is denoted by $\mathbf{x} = \{x_1, x_2, \ldots, x_T\}$, with the corresponding residual vector denoted by $\mathbf{e} = \{e_1, e_2, \ldots, e_T\}$. In this paper, the functional form for the unknown model function $g(x|\theta)$ is represented using a three-layer ANN.

Before a neural network can be used to predict future values of y, it has to be trained using a training data set. The details on how to construct a neural network, how a neural network can be trained, and how to use a trained neural network for forecasting of the target variable can be found in many textbooks and published papers. In particular, we refer readers to Men et al. [8,9] for a description of the theory underpinning neural networks as applied to the ensemble forecasting of wind speed and wind turbine power.

There is a disadvantage in using conventional ANNs for forecasting future values of y. Because there is no distributional assumption for the residuals (or equivalently, there is no conditional distribution for the dependent variable y), the uncertainty quantification of the forecasts of y at various lead times need to be achieved using ensemble methods (e.g., the ensemble ANN method proposed in Men et al. [8,9] and Wang et al. [13]). The members of an ensemble can include potentially the outputs of various types of forecast models (e.g., machine learning approaches based on ANNs, empirical time series modeling approaches based on GARCH, physics-based approaches using NWP models, etc.) The forecasts of y can be obtained by using either a simple average of the forecasts over the component members in the ensemble, or alternatively by using a weighted average of the

individual forecasts where the weights are obtained using some sophisticated approach (e.g., by minimizing an objective loss function such as a quadratic function of the prediction errors as advocated by Men et al. [8,9]).

In this paper, we will represent the uncertainty in the prediction of the output (target) variable y using a probability density of y conditioned on a (given) input variable x, denoted as $f(y|x)$. The functional form of the conditional distribution is usually unknown and, as a consequence, we use a mixture or sum of Gaussian distributions to approximate the conditional distribution of $f(y|x)$. More explicitly, the conditional distribution of y_t given the input vector $\mathbf{x}$ is represented by a mixture of Gaussian distributions as follows:

$$f(y_t|\mathbf{x}) = \sum_{n=1}^{N} p_n(\mathbf{x},t)\phi\big(y_t|\mu_n(\mathbf{x},t),\sigma_n^2(\mathbf{x},t)\big), \qquad t = 1,2,\ldots,T, \tag{2}$$

where $p_n(\mathbf{x},t)$ $(n = 1,2,\ldots,N)$ is the n-th mixing coefficient that defines the weights for each kernel in the mixture and $\phi\big(y_t|\mu_n(\mathbf{x},t),\sigma_n^2(\mathbf{x},t)\big)$ $(n = 1,2,\ldots,N)$ is the n-th kernel in the mixture which is taken herein to be a Gaussian distribution with mean $\mu_n(\mathbf{x},t)$ and variance $\sigma_n^2(\mathbf{x},t)$. In order that $f(y_t|\mathbf{x})$ be a valid representation for the conditional distribution of y_t, the mixing coefficients must be non-negative and verify the constraint $\sum_{n=1}^{N} p_n(\mathbf{x},t) = 1$. With the representation for the conditional distribution of the output variables given in Equation (2), the model likelihood function for the training data set is simply the product of the independent probabilities for the output variables:

$$L(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{T} f(y_t|\mathbf{x}) = \prod_{t=1}^{T}\left(\sum_{n=1}^{N} p_n(\mathbf{x},t)\phi\big(y_t|\mu_n(\mathbf{x},t),\sigma_n^2(\mathbf{x},t)\big)\right). \tag{3}$$

The values for the parameters of the mixture density of Equation (2), namely p_n, μ_n and σ_n^2 $(n = 1,2,\ldots,N)$ can be estimated using a three-layer standard neural network: namely, an input layer, a hidden layer and an output layer. The use of a three-layer ANN architecture is justified by the work of Hornik et al. [16] and White [17] who showed that subject to some regularity conditions, a three-layered ANN can approximate any continuous function of multivariate real variables to any degree of accuracy provided the number of nodes in the hidden layer is sufficiently large. Indeed, seminal work by Bishop [14,15] demonstrated that a mixture density network (MDN) can be used to estimate or learn the probabilistic relationship between the input and output values of an unknown mapping [parametrized in the form of a conditional distribution given by Equation (2)] using a

training data set consisting of a set of examples of the input and output values of the mapping. Nabney [18] describes a software package called NetLab that provides an implementation of a MDN for probabilistic learning. The primary difference between a standard neural network and a MDN resides in the fact that nodes of the output layer are not the values of the output variables y but instead, are the parameters of the mixture density. In consequence, for a MDN, the number of input nodes is the dimensionality N_x of the input variable x and the total number of output nodes is $3N$ where N is the number of kernels present in the mixture density [cf. Equation (2)]. In other words, the total number of output nodes is equal to the number of parameters of the mixture density.

In order to ensure that the mixture density of Equation (2) is physically realizable, it is necessary that the standard deviations σ_n are positive and the mixing coefficients p_n are non-negative with their sum equal to unity. To ensure that these constraints are automatically satisfied, the mixture density parameters p_n, μ_n and σ_n are re-parametrized in terms of α_n, β_n and γ_n as follows. Firstly, the mixing coefficients p_n are represented as

$$p_n(\mathbf{x}, t) = \frac{\exp\big(\alpha_n(\mathbf{x}, t)\big)}{\sum_{n=1}^{N} \exp\big(\alpha_n(\mathbf{x}, t)\big)}, \qquad n = 1, 2, \ldots, N. \tag{4}$$

This re-parametrization ensures that the mixing coefficients are non-negative with the sum of all coefficients equal to one. To ensure the positivity of the standard deviation, σ_n is re-parametrized as

$$\sigma_n(\mathbf{x}, t) = \exp\big(\beta_n(\mathbf{x}, t)\big), \qquad n = 1, 2, \ldots, N. \tag{5}$$

Finally, because there are no constraints on μ_n, we simply define

$$\mu_n(\mathbf{x}, t) = \gamma_n(\mathbf{x}, t), \qquad n = 1, 2, \ldots, N. \tag{6}$$

With this re-parameterization, the parameters α_n, β_n and γ_n become the outputs of the MDN, rather than p_n, μ_n and σ_n.

After the MDN has been properly trained on a training data set, using an expectation-maximization (EM) optimization procedure, we obtain an estimate of the parameter vector $\widehat{\theta}$ of θ [where the components of θ are the parameters that define the mixture density given by Equation (2)]. Given some new input data $\mathbf{x}_{T+1}^{T+d} \equiv \big\{x_{T+1}, x_{T+2}, \ldots, x_{T+d}\big\}$ $(d \geq 1)$, we can utilize the trained MDN to provide (future) predictions for the associated output variables $\widehat{y}_t$ $(t = T+1, T+2, \ldots, T+d)$ using the

mixture conditional density as follows:

$$f\Big(\widehat{y}_t\Big|\mathbf{x}_{T+1}^{T+d}\Big)=\sum_{n=1}^{N}\widehat{p}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\phi\bigg(\widehat{y}_t\Big|\widehat{\mu}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big),\widehat{\sigma_n^2}\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\bigg), \tag{7}$$

with $t=T+1,T+2,\ldots,T+d$. In Equation (7), the hat over a parameter is used to denote an estimate for this parameter obtained from the fully trained MDN. We can interpret the model in Equation (7) as information channels from the new input data to the (future) predictions of the output data (target variable).

With the approximate representation for the mixture density given by Equation (7), the various conditional moments for $\widehat{y}_t$ can be readily obtained. For example, the first conditional moment (or mean) of $\widehat{y}_t$ is given by

$$\begin{aligned}\mathcal{M}_t\Big(\mathbf{x}_{T+1}^{T+d}\Big) &= \mathcal{E}\Big(\widehat{y}_t\Big|\mathbf{x}_{T+1}^{T+d}\Big)\\ &= \sum_{n=1}^{N}\widehat{p}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\widehat{\mu}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big), \qquad t=T+1,T+2,\ldots,T+d,\end{aligned} \tag{8}$$

where $\mathcal{E}(\,\cdot\,)$ denotes mathematical expectation.

The second conditional moment of $\widehat{y}_t$ (about the origin) is similarly given by

$$\mathcal{E}\Big(\widehat{y_t^2}\Big|\mathbf{x}_{T+1}^{T+d}\Big)=\sum_{n=1}^{N}\widehat{p}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\bigg(\widehat{\mu}_n^2\Big(\mathbf{x}_{T+1}^{T+d},t\Big)+\widehat{\sigma_n^2}\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\bigg), \quad t=T+1,T+2,\ldots,T+d. \tag{9}$$

In view of this, the conditional variance of $\widehat{y}_t$ is

$$\begin{aligned}\mathrm{Var}\Big(\widehat{y}_t\Big|\mathbf{x}_{T+1}^{T+d}\Big) &\equiv \mathcal{E}\Big(\widehat{y_t^2}\Big|\mathbf{x}_{T+1}^{T+d}\Big)-\bigg(\mathcal{E}\Big(\widehat{y}_t\Big|\mathbf{x}_{T+1}^{T+d}\Big)\bigg)^2\\ &= \sum_{n=1}^{N}\widehat{p}_n\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\bigg(\widehat{\mu}_n^2\Big(\mathbf{x}_{T+1}^{T+d},t\Big)+\widehat{\sigma_n^2}\Big(\mathbf{x}_{T+1}^{T+d},t\Big)\bigg)-\mathcal{M}_t^2\Big(\mathbf{x}_{T+1}^{T+d}\Big),\end{aligned} \tag{10}$$

for $t=T+1,T+2,\ldots,T+d$. Consequently, at each lead time t, the best forecast for $\widehat{y}_t$ of y can be estimated using the conditional mean $\mathcal{M}_t\big(\mathbf{x}_{T+1}^{T+d}\big)$. The conditional standard deviation in this estimate is given by $s\big(\mathbf{x}_{T+1}^{T+d}\big)\equiv\mathrm{Var}^{1/2}\big(\widehat{y}_t\big|\mathbf{x}_{T+1}^{T+d}\big)$. The uncertainty of the corresponding forecast can be quantified using a confidence interval at some specified confidence level.

Two main sources of uncertainty need to be addressed in the application of MDNs. It is necessary to select the appropriate number of kernels N in the mixture density representation of Equation (2)

and the number of hidden units in the feedforward neural network used to predict the values of the parameters in the mixture density. To address this issue in a semi-automatic manner, we consider the use of an ensemble of MDNs, rather than a single MDN. The component MDNs in the ensemble are identified by the number of nodes in the hidden layer (viz., each member of the ensemble represents a specific neural network architecture). The best prediction of the target vector using the ensemble of MDNs can be obtained as a weighted sum of predictions from the component MDNs. To determine the weight to use for each member MDN of the ensemble, we use the conditional likelihood for each of the individually fitted MDNs. The corresponding weights are selected to be proportional to the value of the likelihood of the fitted MDNs based on the training data set. Once the weights for all the component MDNs have been determined, best forecasts of wind speed and wind turbine power can be obtained. The precision (uncertainty quantification) in these forecasts can be provided in the form of $p\%$ confidence intervals (viz., intervals that will include the true value $p\%$ of the time). Felder [19] introduced mixture density recurrent ANNs for wind power forecasting, but in their methodology only one single ANN was used for the prediction. In this paper, the uncertainty in the neutral network architecture (appropriate number of hidden units) and in the target vector for prediction are characterized using an ensemble of MDNs.

To begin, we start with a MDN with a fixed number of components N in the mixture. Now, suppose we consider an ensemble of M three-layer MDNs, with each MDN member in the ensemble being identified by the number of nodes in the hidden layer. Define by $L_m(\mathbf{y}|\mathbf{x})$ $(m = 1, 2, \ldots, M)$ the likelihood function associated with the m-th component (member) MDN in the ensemble. From Equation (3), the likelihood function for the m-th component MDN is given by

$$L_m(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{T} \left(\sum_{n=1}^{N} \widehat{p}_{mn}(\mathbf{x}, t) \phi\Big(y_t \Big| \widehat{\mu}_{mn}(\mathbf{x}, t), \widehat{\sigma^2_{mn}}(\mathbf{x}, t) \Big) \right), \qquad m = 1, 2, \ldots, M, \tag{11}$$

where the subscript m on all quantities is used to denote that these quantities refer to the m-th MDN in the ensemble.

The multi-step ahead ensemble forecast for the target (output) variable y can be expressed as

$$\begin{aligned} \widetilde{y}_t\big(\mathbf{x}_{T+1}^{T+d}\big) &= \sum_{m=1}^{M} \omega_m \left(\sum_{n=1}^{N} \widehat{p}_{mn}\Big(\mathbf{x}_{T+1}^{T+d}, t\Big) \widehat{\mu}_{mn}\Big(\mathbf{x}_{T+1}^{T+d}, t\Big) \right) \\ &= \sum_{m=1}^{M} \omega_m \mathcal{M}_{mt}\Big(\mathbf{x}_{T+1}^{T+d}\Big), \qquad t = T+1, T+2, \ldots, T+d, \end{aligned} \tag{12}$$

where ω_m are the ensemble weights and $\mathcal{M}_{mt}(\mathbf{x}_{T+1}^{T+d})$ is conditional mean of the target variable obtained from the m-th MDN [and defined earlier in Equation (12)]. To determine the weights ω_m to use for the ensemble model averaging, it is argued that a reasonable estimate for the weight is $\widehat{\omega}_m \propto p\big(m\big|\mathbf{y},\mathbf{x}\big)$ $(m = 1, 2, \ldots, M)$ where $p(m|\mathbf{y},\mathbf{x})$ is the posterior probability for the m-th MDN in the ensemble. According to Bayes' rule, we can determine this probability as follows:

$$p(m|\mathbf{y},\mathbf{x}) = \frac{L_m(\mathbf{y}|\mathbf{x})p(m|\mathbf{x})}{p(\mathbf{y}|\mathbf{x})}, \qquad m = 1, 2, \ldots, M, \tag{13}$$

where $p(m|\mathbf{x})$ is the prior probability for the m-th MDN in the ensemble. Furthermore, $p(\mathbf{y}|\mathbf{x})$, which the conditional probability of the output variables given the input variables, can be seen from Equation (13) to be simply a normalization constant over all the MDN components in the ensemble, viz.

$$p(\mathbf{y}|\mathbf{x}) = \sum_{m=1}^{M} L_m(\mathbf{y}|\mathbf{x})p(m|\mathbf{x}). \tag{14}$$

If the component MDNs in the ensemble are equally probable *a priori* (viz., $p(m|\mathbf{x}) = M^{-1}$ for $m = 1, 2, \ldots, M$), then the posterior probability for the m-th component MDN is simply proportional to the likelihood function for this component MDN, so $p(m|\mathbf{y},\mathbf{x}) \propto L_m(\mathbf{y}|\mathbf{x})$. From this perspective, the weights ω_m can be estimated as follows:

$$\widehat{\omega}_m = \frac{L_m(\mathbf{y}|\mathbf{x})}{\sum_{m=1}^{M} L_m(\mathbf{y}|\mathbf{x})}, \qquad m = 1, 2, \ldots, M. \tag{15}$$

Note that the estimated weights are normalized so that $\sum_{m=1}^{M} \widehat{\omega}_m = 1$.

The standard deviation in the ensemble predictions $\widetilde{y}_t$ for the output variables y_t at the lead times $t = T+1, T+2, \ldots, T+d$, can be determined as follows. Firstly, we calculate the second conditional moment of the predictions about the origin to give

$$\mathcal{E}\Big(\widetilde{y_t^2}\Big|\mathbf{x}_{T+1}^{T+d}\Big) = \sum_{m=1}^{M} \widehat{\omega}_m \mathcal{E}\Big(\widehat{y_{mt}^2}\Big|\mathbf{x}_{T+1}^{T+d}\Big), \qquad t = T+1, T+2, \ldots, T+d, \tag{16}$$

where $\mathcal{E}(\widehat{y_{mt}^2}\big|\mathbf{x}_{T+1}^{T+d})$ is given in Equation (9) when applied to the m-th component MDN of the

ensemble. Secondly, given this information, the conditional variance of $\widetilde{y}_t$ is determined from

$$\begin{aligned}\sigma^2_{\widetilde{y}_t}\left(\mathbf{x}_{T+1}^{T+d}\right) &= \mathrm{Var}\left(\widetilde{y}_t|\mathbf{x}_{T+1}^{T+d}\right)\\ &= \mathcal{E}\left(\widetilde{y}_t^2\middle|\mathbf{x}_{T+1}^{T+d}\right) - \left(\mathcal{E}\left(\widetilde{y}_t\middle|\mathbf{x}_{T+1}^{T+d}\right)\right)^2\\ &= \sum_{m=1}^{M}\widehat{\omega}_m\mathcal{E}\left(\widehat{y_{mt}^2}\middle|\mathbf{x}_{T+1}^{T+d}\right) - \left(\sum_{m=1}^{M}\widehat{\omega}_m\mathcal{M}_{mt}\left(\mathbf{x}_{T+1}^{T+d}\right)\right)^2, \end{aligned} \tag{17}$$

for $t = T+1, T+2, \ldots, T+d$.

The confidence interval at a prescribed confidence level for the ensemble forecasts of the output variable can be obtained as

$$(\widehat{y}_t)_{\mathrm{est}} = \widetilde{y}_t\left(\mathbf{x}_{T+1}^{T+d}\right) \pm \widehat{\sigma}_{\widetilde{y}_t}\left(\mathbf{x}_{T+1}^{T+d}\right)t_{\alpha,v}, \qquad t = T+1, T+2, \ldots, T+d, \tag{18}$$

where $t_{\alpha,v}$ is the α-level critical value of the Student's t-distribution with $v = (M-1)$ degrees of freedom.

Finally, note that the ensemble MDN forecasting of the output variables y is equivalent mathematically to using a single mixture PDF of the form

$$f(y_t|\mathbf{x}) = \sum_{m=1}^{M}\omega_m\sum_{n=1}^{N}p_n(\mathbf{x},t)\phi\big(y_t|\mu_{nm}(\mathbf{x},t),\sigma^2_{nm}(\mathbf{x},t)\big), \tag{19}$$

whose MN mixing coefficients are given by $\omega_m p_n$ with $m = 1, 2, \ldots, M$ and $n = 1, 2, \ldots, N$. It can be seen that $\omega_m p_n$ are valid mixing coefficients owing to the fact that these coefficients are non-negative and their sum must equal unity. From this perspective, the use of an ensemble of MDNs as described above provides a semi-automatic procedure for selecting the appropriate number of hidden units and kernels. This semi-automatic procedure circumvents the trial-and-error procedure that has been used to date for selecting the number of kernels and hidden units of the MDN to use in a particular application (see Bishop [15]).

3. Model selection

For a given data set, there are many appropriate models that can fit the data well and each of these fitted models can be used for the prediction (forecasting) of the target variable. The question that arises is how to select one model that produces the best forecast for the future values of the

target variable. Two common objective functions that are used in the optimization procedure for determination of the optimal parameters for a given model structure are the root-mean-square error (RMSE) given by

$$\text{RMSE} = \left(\frac{1}{T} \sum_{t=1}^{T} \left| e_t \right|^2 \right)^{\frac{1}{2}} = \left(\frac{1}{T} \sum_{t=1}^{T} \left| y_t - g(x_t|\hat{\theta}) \right|^2 \right)^{\frac{1}{2}}, \tag{20}$$

where $\hat{\theta}$ is the estimated parameter from the training process, and the mean absolute error (MAE) given by

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} \left| e_t \right| = \frac{1}{T} \sum_{t=1}^{T} \left| y_t - g(x_t|\hat{\theta}) \right|. \tag{21}$$

Although the RSME and MAE can be applied generally as goodness-of-fit measures for the estimation of θ, these measures can also be used for the assessment of the predictive performance of point (deterministic) forecasts of a target variable.

It is noted that the ensemble MDN approach advocated herein provides probabilistic forecasts of a target variable, rather than simply a point forecast [although the latter is possible by simply using the mean forecast $\widetilde{y}_t\left(\mathbf{x}_{T+1}^{T+d}\right)$ as in Equation (12) and ignoring the remaining information embodied in the predictive distribution of Equation (19)]. For probabilistic forecasting, alternative measures to RMSE and MAE must be used for evaluating probabilistic wind speed and power forecasts. To this end, Gneiting et al. [10] proposed that a proper scoring rule based on the continuous ranked probability score (CRPS) be used as a metric for the rigorous assessment of probabilistic forecasts of wind speed. The CRPS is particularly useful for prediction assessment when ensemble methods are used. Indeed, the CRPS has recently attracted renewed attention for environmental forecasting, for instance, in Grimit and Mass [20], Candille and Talagrand [21], Gneiting et al. [10,22], Wilks [23], Grimit et al. [24], among others.

The CRPS is defined as follows. If $F(y)$ is the predictive cumulative distribution function (CDF) for a target variable y and y^* is the value of y that is realized, then the CRPS is defined as

$$\text{CRPS}(F, y^*) = \int_{-\infty}^{\infty} \left(F(y) - \mathbf{1}(y \geq y^*) \right)^2 dy, \tag{22}$$

where $\mathbf{1}(y > y^*)$ is an indicator function that takes the value 1 when $y \geq y^*$ and the value 0, otherwise. Gneiting and Raftery [25] showed that

$$\text{CRPS}(F, y^*) = \mathcal{E}_F|Y - y^*| - \frac{1}{2}\mathcal{E}_F|Y - Y'|, \tag{23}$$

where Y and Y' are two independent realizations of the target variable from the predictive distribution F. From the latter equation, it is readily seen that the CRPS generalizes the mean absolute error, and can be reported in the same physical units as the target variable. Indeed, if F reduces to a point (deterministic) forecast, the CRPS reduces to the MAE [cf. Equation (21)]. Obviously, the CRPS provides a direct method for comparison of probabilistic forecasts, with smaller values for the CRPS corresponding to the better probabilistic forecast.

4. An application to wind speed and power forecasting

4.1. Wind speed and power data

We present an application of our ensemble MDN approach to wind speed and wind turbine power forecasting. The data for this application was obtained from a wind turbine farm in Taiwan. Measurements of the wind speed at hub height and the power generated by a particular wind turbine (which is referred to henceforth as WT#6) in this wind farm are used for validation of our proposed approach.

The input variables x for the MDN are the predicted values for the wind speed obtained from a numerical weather prediction model. To this purpose, predictions of the wind speed at the location of WT#6 in the wind park in Taiwan were obtained using the Weather Research and Forecasting (WRF) model version 3.6 compiled with the Advanced Research WRF (ARW) dynamical core. This model was used to predict the wind speed in the wind farm for the time period from 3–11 January, 2013. Towards this objective, WRF was initialized at 00:00 UTC each day and used a 48 h simulation period. The first 6 h of the simulation period were used as the spinup for the simulation and, as a consequence, was excluded from the analysis. The simulations were initialized and driven by the National Center for Environmental Prediction/Environmental Modeling Center's global forecasting system final analysis data set with a 6 h temporal resolution and a 1-degree horizontal spatial resolution.

Four simulation domains were used to cover the appropriate area of interest and to ensure that the prescribed resolution requirements for the simulation are satisfied. The number of grid points in the coarse domain, centered at 25° N latitude and 115° E longitude, was 222×128 with a grid cell size of 45 km. The coarse domain covers China, most parts of Southeast Asia and the East China Sea. In addition to the coarse domain, three progressively finer domains, which were nested within each other in order to downscale the spatial resolution from 45 km to 1.67 km, were used for

the simulations. The finest domain consists of 151×244 grid points and was situated so that this domain covered the north coast of Taiwan (where the wind park was located) and a large part of the East China Sea with high spatial resolution. Vertically, the model domain was discretized using 40 layers which extended from the ground surface to a height of 50 hPa.

In the simulations, we used the Goddard Cumulus Ensemble (GCE) scheme described in Tao and Simpson [26] for the convective microphysics parameterization, the Kain-Fritsch scheme (Kain and Fritsch [27]) for the cumulus parameterization, the Yonsei University scheme (Hong and Lim [28]) for the planetary boundary-layer parameterization, the revised MM5 Monin-Obukhov scheme (Monin and Obukhov [29]) for the representation of the surface layer in the atmosphere, and the Noah-MP land-surface scheme (Chen and Dudhia [30]) for the parameterization of the land surface model. The Rapid Radiative Transfer Model (RRTM) scheme of Mlawer et al. [31] was selected for the determination of the long-wave radiation and the Goddard shortwave scheme of Chou and Suarez [32] was used for the calculation of the short-wave radiation. The dynamical evolution of the meteorological fields was simulated by the WRF model and the wind speed was output every 30 minutes at the four grid points surrounding wind turbine WT#6 at a height of 45 m. A bilinear interpolation (BI) was then used to interpolate these wind speeds to the location of WT#6 at the hub height of 45 m.

4.2. Predictive performance

The wind speeds at the location of WT#6 were measured every 30 minutes for a period of 8 consecutive days. In consequence, there are 384 observations of the wind speed in total. Correspondingly, we have 384 observations of the power generated by the wind turbine WT#6. The exogenous information used as the input to the neural network are the modelled wind speeds obtained from the NWP model and bilinearly interpolated to provide the modelled wind speed at hub height at the location of WT#6. Each member of the ensemble of MDN models was trained using the first 240 observations corresponding to the first 5 days of data, and the remaining 144 observations corresponding to the last 3 days of data were used for the forecast assessment. The forecast was for the half-hourly averaged wind speed and wind turbine power for a forecast lead time of up to 72 h.

It is useful to compare the numerical predictions of the NWP modelled (BI) wind speeds at WT#6 with the corresponding measured wind speeds at this location. To this purpose, Fig. 1 exhibits the predicted wind speeds from the NWP model plotted against the measured wind speeds at the hub height of the turbine. Note that there is a significant bias in the prediction of the wind

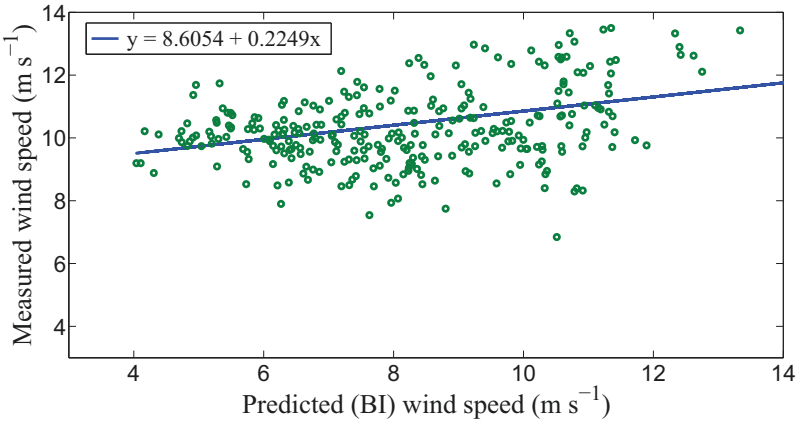

Figure 1: Scatter plot of NWP predicted (BI) wind speed versus the measured wind speed for wind turbine WT#6. The solid line shows the linear regression of the predicted and measured wind speeds (least-squares fit of the data points to a straight-line model). The fit has a coefficient of determination of $r^2 = 0.1708$.

speeds at WT#6 (which from the figure is seen to be about 9 m s^{-1} on average). To characterize the relationship between the predicted and measured wind speeds, a linear regression of the predicted (BI) wind speed on the measured wind speed is superimposed on the scatter plot in Fig. 1. The adjusted r^2 value (or coefficient of determination) for this fit of 0.1708 differs significantly from unity, indicating that the fitted regression model is not appropriate for describing the relationship between the measured wind speeds and the NWP modelled wind speeds. In consequence, the relationship between the measured and modelled wind speeds will be represented non-parametrically using either an ANN or MDN.

In Fig. 2, we exhibit the measured power curve relationship for WT#6 (viz., the relationship between the measured wind speed and the measured wind power for WT#6). A perusal of the measured power curve for WT#6 shows that when the wind speed exceeds 14 m s^{-1} (rated output wind speed), the wind power output is limited to 650 W (rated output power for the wind turbine).

We trained an ensemble of MDNs with the number of kernels N in each MDN set to three. The ensemble of feedforward three-layered MDNs involved varying the number of nodes in the hidden layer from 5 to 204 inclusive (namely, the ensemble consists of $M = 200$ model structures for the MDN, each structure differing in the number of hidden units used to represent the training data set). It should be emphasized that in accordance to Equation (19), the ensemble can be considered to correspond to the flexibility in the representation that can be obtained potentially from a mixture or sum of Gaussian distributions with $MN = 600$ kernels.

In Fig. 3, we compare the measured wind speeds at hub height for WT#6 with the predicted wind

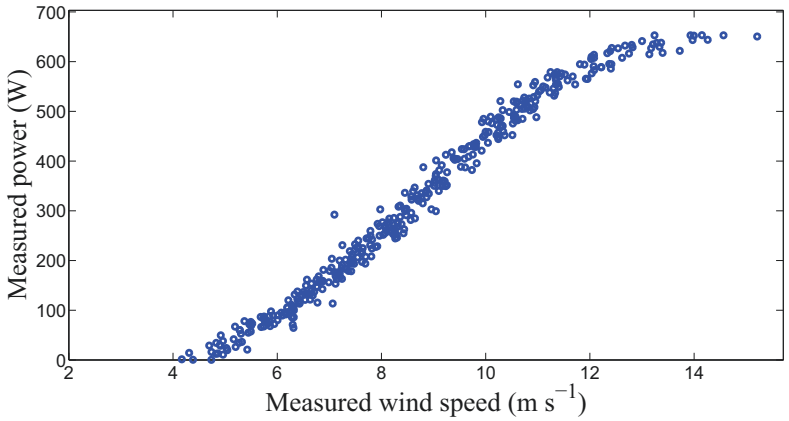

Figure 2: Measured power curve for the wind turbine WT#6.

speeds obtained from the NWP model and the forecasted wind speeds obtained using the ensemble MDNs. A perusal of this figure shows that the forecasted wind speeds obtained from the ensemble MDNs are in better conformance with the measured wind speeds than the predicted wind speeds obtained from the NWP model. Furthermore, this figure displays the 95% confidence intervals for the forecasted wind speeds obtained from the ensemble MDNs. A comparison between the measured and forecasted wind speeds using an ensemble nonlinear autoregressive exogenous (NARX) ANN [8] is shown in Fig. 4. Taken together, Figs. 3 and 4 provide a qualitative assessment of the predictive performance of the various forecast techniques (e.g., NWP model, ensemble MDN and ensemble NARX ANN). However, it is more informative to consider quantitative measures for the assessment of the point and probabilistic forecasts.

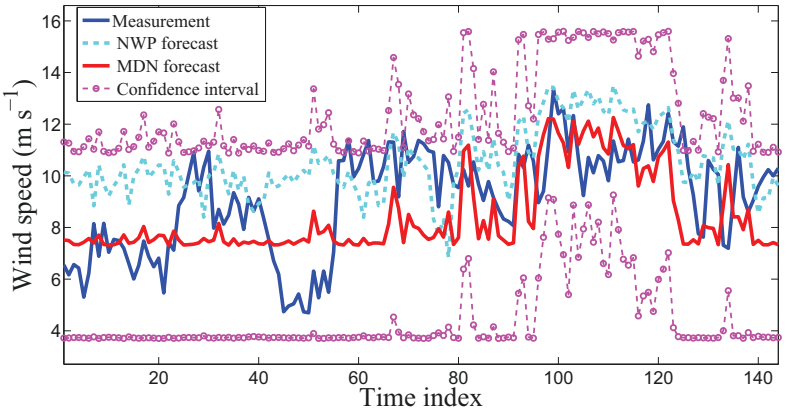

Figure 3: Comparison of the wind speed forecast for WT#6 obtained by the NWP model and the ensemble MDN model. The 95% confidence interval for the ensemble MDN forecast of the wind speed is shown.

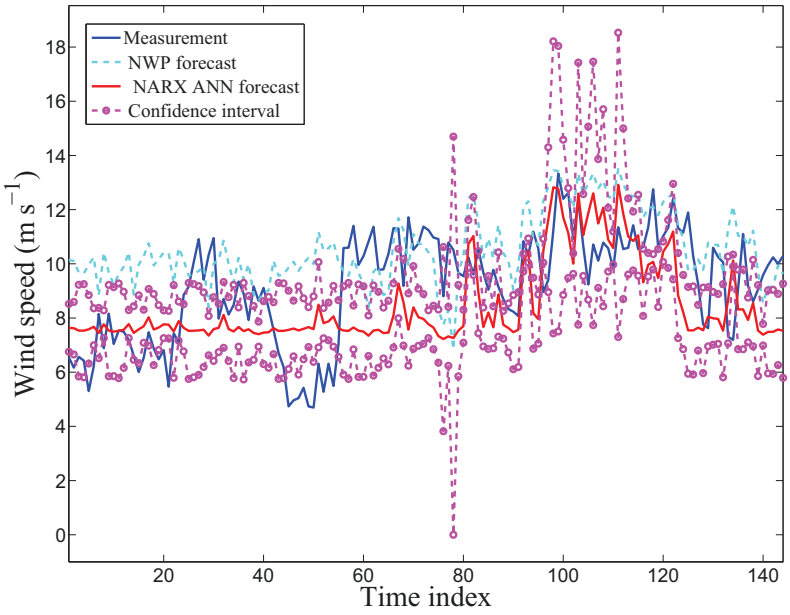

Figure 4: Comparison of the wind speed forecast for WT#6 obtained by the NWP model and the ensemble NARX ANN model of Men et al. [8]. The 95% confidence interval for the ensemble NARX ANN forecast of the wind speed is shown.

Table 1: Wind speed forecast assessment for various cases for WT#6.

Criterion	Persistence	NWP	Ensemble NARX ANN	Ensemble MDN	GARCH(p, q)	ARMA(p, q)
RMSE	3.8526	2.4096	1.9837	**1.9688**	2.0847	2.0061
MAE	3.4006	1.9225	1.6862	**1.6665**	1.8017	1.6965
CRPS	n.a.	n.a.	1.4177	**1.3702**	1.7059	1.8860

To this purpose, Table 1 summarizes the values of the RMSE, MAE and CRPS measures for various forecasts (both point and probabilistic). These cases include the forecasts provided by a persistence model (which uses the current wind speed to predict the value of the future wind speeds), by a NWP (or, physics-based) model, by an ensemble MDN model, by an ensemble NARX ANN model, by an ensemble GARCH(p, q) model, and by an ensemble autoregressive moving average ARMA(p, q) model. In the GARCH(p, q) model, p is the order of the GARCH terms (or, the extent in which the current volatility depends on the previous observations) and q is the order of the ARCH terms (or, the extent in which the current volatility depends on the previous volatilities). Similarly, in the simpler ARMA(p, q) model, p and q correspond to the order of the autoregressive and moving average terms in the model, respectively. In our current application of the GARCH(p, q) and ARMA(p, q) models, the orders p and q are restricted to be less than 4.

With respect to the RMSE, MAE and CRPS measures, it can be seen from a perusal of Table 1 that an ensemble MDN model forecast outperformed all other forecasts methodologies tested for both point forecasts as well as for probabilistic forecasts of the wind speed. In particular, as a determinstic forecast, the ensemble MDN forecast had a MAE that is 104%, 15.4%, 1.2%, 8.1% and 1.8% less than the persistence, NWP, ensemble NARX ANN, GARCH(p, q) and ARMA(p, q) forecasts, respectively, for the 72-h ahead wind speed. Based on the CRPS, it is seen that the ensemble MDN model probabilistic forecasts of the 72-h ahead wind speed has a value for this score that is 3.5%, 25.4% and 37.6% less than the ensemble NARX, GARCH(p, q) and ARMA(p, q) forecasts, respectively.

In summary, it is noted that the ensemble MDN point forecast is marginally better than that provided by the ensemble NARX ANN, and that both these point forecasts are moderately better that the point forecast provided by the NWP model. Not surprisingly, the ensemble neural network (MDN or ANN) and the NWP point forecasts for the wind speed are significantly better than that provided by the simple persistence (no-change) model. This is not surprising owing to the long forecast lead time (72 h) considered in this example. Finally, probabilistic forecasts provided by the ensemble MDN model outperform all others tested, although it should be noted that the ensemble MDN probabilistic forecasts had only a slightly lower CRPS than the ensemble NARC ANN probabilistic forecasts.

The objective functions used for training the MDNs are highly nonlinear multimodal functions of the network parameters, with the result that optimization methods may only provide a sub-optimal

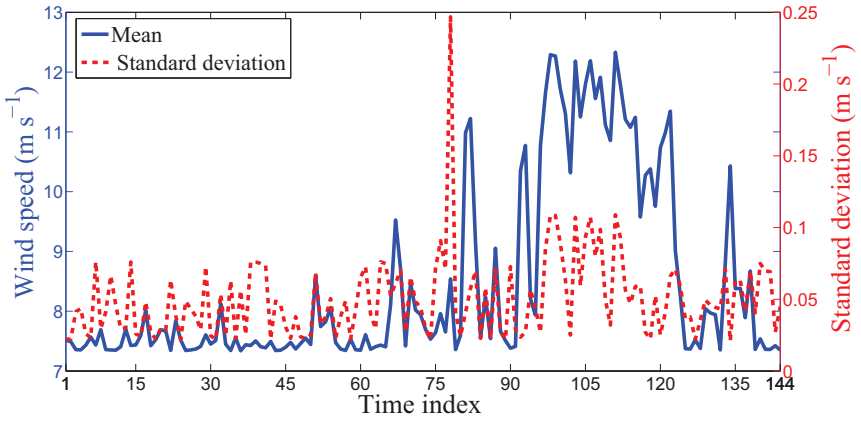

Figure 5: Uncertainty arising from the initialization of the MDN parameters in the application of the ensemble MDN methodology. The mean and standard deviation in the forecasted wind speeds were obtained from 50 independent ensemble MDNs.

solution (viz., a solution corresponding to one of the local minima of the objective function). Furthermore, the quality of the final solution may depend critically on the initial values used for the network parameters. For this reason, there may be some significant uncertainty arising from the initialization of the network parameters since the quality of the solution obtained in the neural network training may depend strongly on this initialization. We investigated this source of uncertainty in our wind speed predictions by executing the ensemble MDN approach on the same training data set 50 different times. In this ensemble of ensemble MDNs, the overall mean and standard deviation of the forecasted wind speeds obtained from the 50 ensembles were computed. A biplot summarizing this analysis is shown in Fig. 5. The left axis in the plot displays the mean of the forecasted wind speed obtained from the 50 ensembles and the right axis in the plot shows the standard deviation in the forecasted wind speed computed over the 50 ensembles. Note that the standard deviation in the forecasted wind speeds is approximately one percent of the overall mean in the forecasted wind speeds (taken over the 50 ensemble MDNs). This result implies that the uncertainty arising from the initialization of the MDN parameters is small, and its contribution to the uncertainty in the wind speed forecasts is small (viz., essentially negligible).

Figure 6 compares the ensemble MDN forecast of the wind power generated by WT#6 with the measured wind power for this turbine. In addition, this figure shows the 95% confidence interval for the predicted wind power obtained from the ensemble MDN. Similarly, Fig. 7 shows a comparison between the measured wind power and the predicted wind power based on the ensemble NARX ANN model. Finally, Table 2 summarizes the values for the RMSE, MAE, and CRPS measures for the

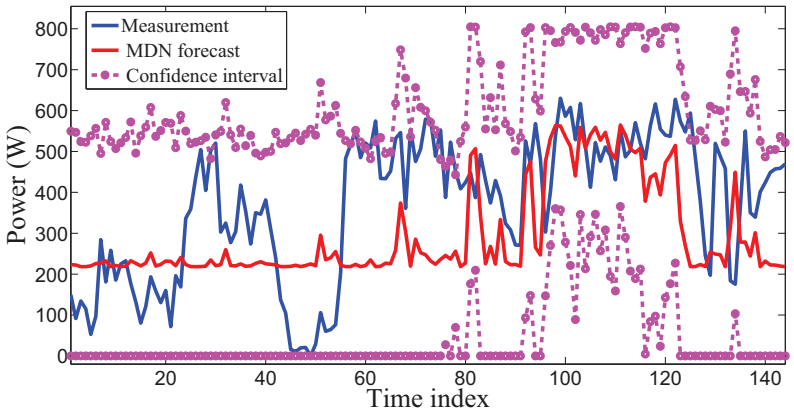

Figure 6: Comparison of the wind power forecast for WT#6 obtained by the ensemble MDN model. The 95% confidence interval for the ensemble MDN forecast of the wind power is shown.

wind turbine power forecast obtained from a persistence model, an ensemble NARX ANN model, an ensemble MDN model, an ensemble GARCH (p, q) model and an ensemble ARMA(p, q) model. As in the case for wind speed forecasting, it is seen that the ensemble MDN model provides the best forecast (both deterministic and probabilistic) for the wind turbine power based on the measures of assessment embodied in RMSE, MAE and CRPS. This forecast is seen to be slightly better than that provided by the ensemble NARX ANN model (point and probabilistic), and both ensemble neural network models yield significantly better point forecasts than that given by the persistence model. This should not be too surprising given the long forecast lead time (72 h). Finally, the ensemble MDN predictive CDFs exhibited significantly lower values for the CRPS than the predictive CDFs provided by the ensemble GARCH and ARMA models.

Table 2: Wind power forecast assessment for various cases for WT#6.

Criterion	Persistence	Ensemble NARX ANN	Ensemble MDN	GARCH(p, q)	ARMA(p, q)
RMSE	339.23	176.72	**174.38**	170.10	187.93
MAE	298.29	149.52	**146.50**	147.41	163.85
CRPS	n.a.	136.77	**133.74**	157.45	159.16

5. Conclusion

An ensemble approach based on mixture density neural networks has been applied to the point and probabilistic forecasting of wind speed and wind turbine power. The ensemble MDN model

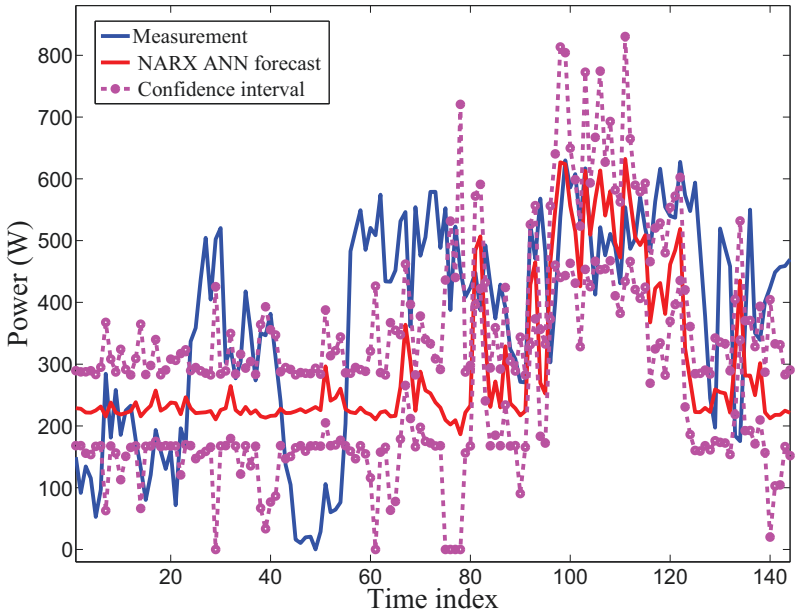

Figure 7: Comparison of the wind power forecast for WT#6 obtained by the ensemble NARX ANN model of Men et al. [8]. The 95% confidence interval for the ensemble NARX ANN forecast of the wind power is shown.

provides a semi-automatic procedure for selecting the number of required kernels and hidden units of the mixture density network that is consistent with the number of samples in the training data set without requiring a tedious trial-and-error procedure. An application of the proposed methodology to some measured data (wind speed and power) obtained from a wind turbine in a wind park in Taiwan demonstrated that the methodology works well. The wind speed and power forecasts were generally better than those obtained from an ensemble ANN model based on a NARX structure for both deterministic and probabilistic forecasts of the wind speed and power. In particular, probabilistic forecasts of these two quantities using an ensemble MDN model performed better than an ensemble GARCH or ARMA models (as measured by the CRPS value). Finally, the ensemble MDN model provides an uncertainty quantification for the wind speed and wind turbine power forecasts in the form of a confidence interval at a prescribed confidence level.

References

[1] Engle RF. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 1982;50:987–1007.

[2] Bollerslev T. Generalized autoregressive conditional heteroskedasticity. J Econometrics 1986;31:307–

414 27.

415 [3] Kretzschmar R, Eckert P, Cattani D, Eggimann F. Neural network classifiers for local wind
416 prediction. J Appl Meteorol 2004;43:727–38.

417 [4] Brown BG, Katz RW, Murphy AH. Time series models to simulate and forecast wind speed and
418 wind power. J Clim Appl Meteorol 1984;23:1184–95.

419 [5] Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production fore-
420 casts. Wind Energy 2004;7:119–32.

421 [6] Bremnes JB. Probabilistic wind power forecasts using local quantile regression. Wind Energy
422 2004;7:47–54.

423 [7] Roulston MS, Kaplan DT, Hardenberg J, Smith LA. Using medium-range weather forecasts to
424 improve the value of wind energy production. Renew Energ 2003;28:585–602.

425 [8] Men Z, Yee E, Lien F-S, Yang Z, Liu Y. Ensemble nonlinear autoregressive exogenous artifi-
426 cial neural networks for short-term wind speed and power forecasting. International Scholarly
427 Research Notices 2014;Article ID 972580:16 pp.

428 [9] Men Z, Yee E, Lien F-S, Ji H, Liu Y. Bootstrapped multi-model neural-network super-ensembles
429 for wind speed and power forecasting. Energy and Power Engineering 2014;6:340–48.

430 [10] Gneiting T, Larson K, Westrick K, Genton MG, Aldrick E. Calibrated probabilistic forecasting
431 at the Stateline Wind Energy Center: The regime-switching space-time method. J Am Stat
432 Assoc 2006;101:968–79.

433 [11] Giebel G, Brownsword R, Kariniotakis G. The state-of-the-art in short-term prediction of wind
434 power. Deliverable Report D1.1, Project Anemos, available at http://anemos.cma.fr/download/-
435 ANEMOS_D1.1_StateOfTheArt_v1.1.pdf.

436 [12] Foley AF, Leahy PG, Marvugliaand A, McKeogh EJ. Current methods and advances in fore-
437 casting of wind power generation. Renew Energ 2012;37:1–8.

438 [13] Wang D, Wang F, Wang X. Multi-step-ahead combination forecasting of wind speed using
439 artificial neural networks. Research Journal of Applied Sciences, Engineering and Technology
440 2003;5:5443–49.

441 [14] Bishop CM. Mixture density networks. Technical Report NCRG/94/004, Neural Computing
442 Research Group, Aston University; 1994.

443 [15] Bishop, CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.

444 [16] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approxima-

445 tors. Neural Networks 1989;2:359–66.
446 [17] White, H. Connectionist nonparametric regression: Multilayer feedforward networks can learn
447 arbitrary mappings. Neural Networks 1990;3:535–49.
448 [18] Nabney, IT. NetLab algorithm for pattern recognition. London: Springer; 2004.
449 [19] Felder, M, Kaifel A, Graves A. Wind power prediction using mixture density recurrent neural
450 networks. European Wind Energy Conference and Exhibition, Poster P0.153, Warsaw, Poland:
451 20–23 April 2010.
452 [20] Grimit EP, Mass CF. Initial results of a mesoscale short-range ensemble system over the Pacific
453 Northwest. Weather and Forecasting 2002;17:192–205.
454 [21] Candille G, Talagrand O. Evaluation of probabilistic prediction systems for a scalar variable. Q
455 J R Meteorol Soc 2005;131:2131-50.
456 [22] Gneiting T, Raftery AE, Westveld AH, Goldman T. Calibrated probabilistic forecasting using en-
457 semble model output statistics and minimum CRPS estimation. Mon Weather Rev 2005;33:1098–
458 1118.
459 [23] Wilks DS. Statistical methods in the atmospheric sciences. 2nd ed. Amsterdam: Elsevier
460 Academic Press; 2006.
461 [24] Grimit EP, Gneiting T, Berrocal VJ, Johnson NA. The continuous ranked probability score for
462 circular variables and its application to mesoscale forecast ensemble verification. Q J R Meteorol
463 Soc 2006;132:1–17.
464 [25] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Am Stat
465 Assoc 2007;102:359–78.
466 [26] Tao WK, Simpson J. The Goddard cumulus ensemble model. Part I: Model description. Terr
467 Atmos Oceanic Sci 1993;4:35–72.
468 [27] Kain JS, Fritsch JM. Convective parameterization for mesoscale models: The Kain-Fritcsh
469 scheme, The representation of cumulus convection in numerical models. Meteor Monogr
470 1993;24:165–70.
471 [28] Hong SY, Lim JOJ. The WRF single-moment 6-class microphysics scheme (WSM6). J Korean
472 Meteor Soc 2006;42:129–51.
473 [29] Monin AS, Obukhov AM. Basic laws of turbulent mixing in the surface layer of the atmosphere.
474 Contrib Geophys Inst (Acad Sci USSR) 1954;151:163–87 (in Russian).
475 [30] Chen F, Dudhia J. Coupling an advanced land-surface/hydrology model with the Penn State/NCAR

476 MM5 modeling system. Part I: Model description and implementation. Mon Weather Rev
477 2001;129:569–85.
478 [31] Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomo-
479 geneous atmosphere: RRTM, a validated correlated-k model for the longwave. J Geophys Res
480 1997;102(D14):16663–82.
481 [32] Chou MD, Suarez MJ. An efficient thermal infrared radiation parameterization for use in general
482 circulation models. NASA Tech Memo 1994;3:85 pp.