

Ensemble Learning of Numerical Weather Prediction for Improved Wind Ramp Forecasting

Xiaomei Chen, Jie Zhao, and Miao He

Dept. ECE, Texas Tech University, Lubbock, TX 79409

Email: {Xiaomei.Chen, Jie.Zhao, and Miao.He}@ttu.edu

Abstract—Numerical weather predictions used for wind power forecasting might not be updated in a timely manner in practice, due to its high computational complexity and complicated post-processing. Thus, the accuracy of wind power forecasts could be significantly compromised especially during wind ramp events. This paper presents an innovative method for improving regional wind power ramp forecasting through ensemble learning of numerical weather prediction models, by using real-time weather measurements as the supervisory data. The numerical weather prediction models are combined to minimize the discrepancy between the forecast values and the real-time measurements in the trend of wind ramps, and the weights of the linear combination are calculated through gradient boosting. The proposed method is non-intrusive and could be efficiently carried out online. The proposed method is evaluated on historical ERCOT wind power ramp events, and compared with existing ensemble aggregation method using simple averaging. The results reveal the effectiveness of the proposed method for improving wind power forecasting during wind ramp events.

Index Terms—Ensemble learning, numerical weather prediction, wind ramp forecasting

I. INTRODUCTION

Grid integration of wind power has been benefited by enhanced short-term wind power forecasting that are based on numerical weather prediction (NWP) [1]. NWP uses mathematical models of the atmosphere to predict future weather conditions [2] and have been widely incorporated as the engine for wind power forecasting systems [3]. These mathematical models are comprised of partial differential equations that govern the physical principles of the atmosphere, which are solved numerically with given initial conditions. Then, the output of NWP is post-processed by model output statistics [4], for which statistical methods are applied to the output of NWP models so that they match the surface observations at lower layers. When applied to short-term wind power forecasting, NWP faces two key technical challenges. First, numerically solving partial differential equations can be very computationally intensive [5], and meanwhile, the error caused by the disparity between the numerical solution and the exact solution of the partial differential equations could accumulate over the rolling horizon [6]. Second, these partial differential equations are typically conditioned or parameterized by exogenous physical processes (e.g., solar radiation, terrain condition, etc.) [7], which introduces a variety of inherent uncertainty to the models and prediction results. As a practical solution, ensemble forecasting has been utilized to gauge the confidence of forecast by accounting for the stochastic nature and inherent

uncertainty of atmospheric processes [8]. Basically, ensemble forecasting utilizes multiple NWP models and/or by varying the physical parametrization or initial conditions of individual models. Particularly, probabilistic wind power forecasts could also be produced from ensemble forecasts [6].

Due to the high computational burden, NWP produces output sporadically. For example, the major NWP systems in practice, including the Global Forecast System (GFS) and North American Mesoscale Forecast System (NAM), are refreshed every six hours [9], and the Weather Research and Forecasting Model (WRF) is run 8 times a day [10]. Most recently, effort has been made toward incorporating the hourly-updated Rapid Refresh (RAP) NWP for wind forecasting through the Wind Forecast Improvement Project II [11]; however, the extreme computational expense of numerical experiments prevents its immediate application to wind power forecasting. Consequently, during the time period of a few hours when the NWP results are yet to be refreshed, the actual wind speed can deviate dramatically from the forecasts, especially for wind ramp events. Then, the resulting wind power forecasts could have significant error. To deal with this challenging issue, this paper investigates post-processing methods of ensemble NWP for improved wind power forecast.

A literature survey on short-term wind power prediction is given by reference [12], which covers the state-of-the-art NWP models that have been applied to wind power forecasting. Wind power forecasting that utilizes ensemble NWP and prognosis of forecast uncertainty is proposed in reference [13]. An approach for constructing probabilistic wind power forecasting from NWP models, together with calibration and smoothing involving the use of statistical time series models, is proposed in reference [14]. The factor of diurnal cycle and its impact on wind power forecasting based on both deterministic and ensemble NWP systems are revealed in reference [15]. Wind power re-forecast using NWP for generating past forecast data that could be combined with historical observation data for statistical model building and calibration is presented in reference [16], and case studies using major NWP systems are conducted. The statistical features of NWP are explored in reference [17], based on which an autoregressive integrated moving average model is constructed to adjust the prediction output for improved wind power forecast. A novel approach for mining the bad data of NWP that constructing improved wind power forecasting model is proposed by reference [18]. Reference [19] utilizes three-dimensional convolutional neural

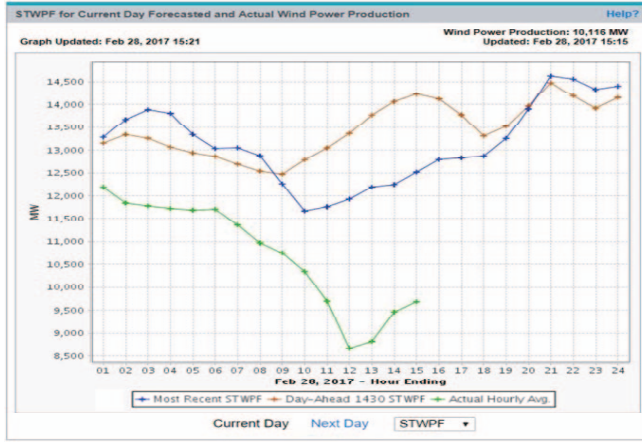


Fig. 1. ERCOT hourly wind power, 15 PM, Feb 26th, 2017.

networks to extract the spatio-temporal features from the output data of NWP for wind farm power forecasting. Different from existing work, this paper is focused on the critical time window (1-3 hours ahead of impending wind power ramps) while refreshed NWP results are not available, and utilizes real-time weather measurements to improve the forecast of ensemble NWP models.

The rest of the paper is organized as follows. A motivating case and the NWP models used in this study are introduced in Section II. Section III presents the proposed approach for improve the forecast of ensemble NWP models. Numerical experiments using real-world data are presented in Section IV. Finally, conclusions are given in Section V.

II. NWP FOR WIND POWER FORECASTING

A. Wind Power Forecasting

The ERCOT's short-term wind power forecast (STWPF) is produced hourly for an episode of rolling 168 hours for the wind production potential of each wind generation resource in ERCOT. The primary service vendor for ERCOT's wind power forecasting is UL/AWS Truepower [20], which utilizes the Mesoscale Atmospheric Simulation System (MASS) [21] as NWP. Hourly forecast for each wind generation resource is delivered to individual wind generation resource 15 minutes after the hour, and the aggregate forecast is available for public access [22]. These forecasts are used by ERCOT for day-ahead and intra-day reliability unit commitment as well as by wind power producers for plant scheduling and market operations. Due to the accuracy or refreshing rate of NWP, hourly wind power forecasting of ERCOT might have significant errors. Such an instance is illustrated in the forecasting snapshot [22] of Fig. 1, where the day-ahead and the most recent STWPF (the red and blue lines in Fig. 1, restively) deviated significantly from the realized wind power (the green line in Fig. 1) during a large down ramp event. It is observed that during the time period from 6 AM-10 AM, the most recent hourly forecasts captured the trend of the down ramp, but failed to quantify the magnitude correctly. More importantly,

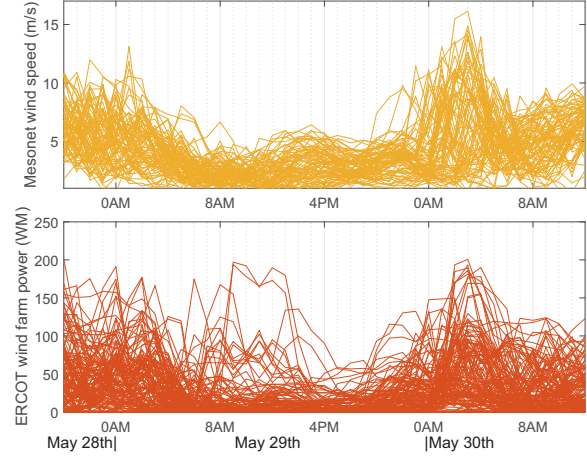


Fig. 2. Mesonet wind speed and ERCOT hourly wind power (May, 2015).

during the time period from 10 AM-12 PM, the realized wind power remained in the down ramp, while the most recent hourly forecasts began to ramp up. On key observation is that if this evidence of down ramp could be incorporated to calibrate the NWP-based forecast, the huge forecasting error (amounting to 3.3 GW) at 12 PM could be mitigated. Motivated by this event, this paper is devoted to developing methods that utilize real-time observation data to calibrate and improve wind power forecasting from an ensemble of NWP models, especially for wind power ramp events.

B. Real-time Weather Measurement

As could be seen from Fig. 4(a), a majority of ERCOT's wind farms are located in the West Texas and Panhandle regions. These regions are covered and monitored by the Mesonet system [23] comprised of over 130 observation stations. The Mesonet stations collect measurement data on weather observations, including wind speed, direction, temperature, solar radiation, humidity, etc. The stations take measurements every 3 seconds, and send 1-minute average values to data center. Figure 2 illustrates the Mesonet measurement data on hourly average wind speed versus the ERCOT farm-level hourly wind power production, plotted for 86 Mesonet sites and 135 ERCOT wind farms that have no missing data during that time period. It could be seen that the 'envelop' formed by the wind speed time series has a similar shape with that of the ERCOT wind farm power production time series; however, it leads the wind power down ramp occurred from 2 AM to 6 AM, May 29th, by 4 hours, and the wind power up ramp occurred from 0 AM to 3 AM, May 30th, by 3 hours. This is because the Mesonet stations cover an extended geographical region, and thus the ramp of observed wind speed could lead that in the wind power production within that region.

C. Ensemble NWP

The ensemble NWP used in this study is from the real-time weather prediction system [24], which contains a 42-member ensemble forecast system and utilizes the Advanced Research WRF (WRF-ARW) model [25]. The NWP models

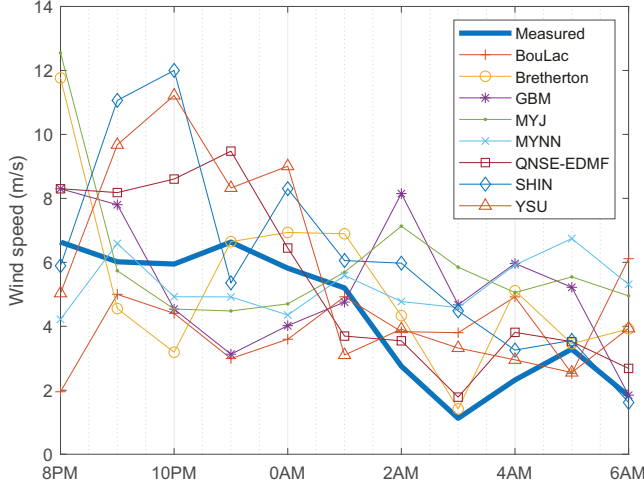


Fig. 3. Wind speed measurements and NWP data.

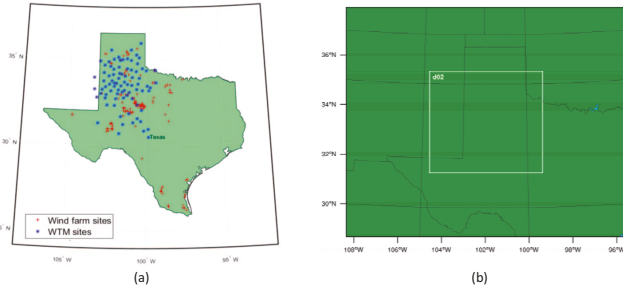


Fig. 4. (a) Mesonet sites and ERCOT wind farms and (b) WRF simulation domain

utilize 12 km-by-12 km grid cells encompassing the southwest U.S. and 4 km-by-4 km grid cells over Texas and portions of surrounding states for two nested domains (see Fig. 4(b)) possessing 38 vertical levels. The ensemble comprised of eight physical schemes are both data assimilation and forecasting systems. The initial conditions are obtained through an ensemble Kalman filter assimilation procedure [24], and more ensemble members are created by varying the initial conditions from these eight physical schemes. The output of the ensemble forecast system is updated every three hours.

For wind power down ramp event on the May 29th, 2015, the wind speed measurement of a selected Mesonet station is compared with the ensemble forecasts, as illustrated in Fig. 3. The corresponding WRF-ARW NWP is run by using re-analysis data available at 0300 UTC for the eight physical schemes, and is then interpolated for 120 m (which is the typical wind turbine hub height). The Mesonet wind speed measurements are collected at an elevation of 33ft (10m). Note that the NWP for 10m is unattainable as the lowest vertical layer is above 10m, and also, the NWP values for 120m are generally higher than the measured ones at 10m, as can be seen from Fig. 3. One key observation from Fig. 3 is that for the down ramp time period, i.e., from 0 AM-4 AM, the measured wind speed drops dramatically; however, not all the NWP values are following the same trend. Specifically, four among the eight schemes, ‘SHIN’, ‘YSU’, ‘QNSE-EDMF’,

and ‘Breherton’, correctly predict the trend (dropping); in sharp contrast, for some other schemes, the forecasted values are even climbing. In this case, if all eight schemes are used with equal weights to produce a point forecast, a significant forecasting error could be produced. However, if any side information or evidence on the credibility of the eight schemes (e.g., by inspecting how their predicted wind speed matches the real-time Mesonet measurements), the significant forecast error could be reduced by assigning different weights to NWP models. The above observation provides significant insights on utilizing real-time Mesonet measurements to weigh the schemes in an ensemble NWP, built on which the proposed approach is developed as described in what follows.

III. PROPOSED APPROACH

The proposed approach first scores the NWP models based on how well their forecasts match the Mesonet wind speed measurements, and then additively combines the NWP models.

A. Scoring NWP Models

How well an NWP model matches the wind speed measurements could be easily measured by the model’s forecasting error. However, one tricky issue arises that the wind speed measurements are collected at 33ft (10m) while the output of the NWP does not contain data for 33ft (10m). To circumvent this situation, the sample correlation coefficient between the NWP and the wind speed measurements could be utilized instead; this is plausible since in the example in Fig. 3, it has been seen that correctly predicting trend is of top priority for wind ramps. Figure 5 shows the centered and normalized wind speed data of two NWP models, in comparison to the wind speed measurements for one Mesonet site during a 4-hour time window (10PM, May 28th - 1AM, May 29th, 2015). Note that the 4-hour time window is immediately ahead of the large down ramp event in Fig. 2. It can be seen that the model ‘BouLac’ predicts the wrong trend, resulting in a low correlation coefficient of -0.884, while the model ‘QNSE-EDMF’ matches the measured wind speed very well with a correlation coefficient as high as 0.935. Indeed, the model ‘QNSE-EDMF’ predicts correctly the trend in the wind speed for the following down ramp, as seen from Fig. 3. With this insight, the NWP models could be scored as follows.

Specifically, let $\mathbf{w}_{nt} = (w_{n(t-T)}, \dots, w_{n(t-1)})^T$ be the wind speed measurements at the n -th ($n=1, \dots, N$) Mesonet site within a time window of size T that is immediately ahead of the forecasting time instant t . Let u_{nt} and σ_{nt} denote the sample-based mean and standard deviation of the wind speed measurements of the n -th Mesonet site. Then, the centered and normalized wind speed measurements are given by:

$$\mathbf{y}_{nt} = \frac{\mathbf{w}_{nt} - u_{nt}}{\sqrt{T - 1}\sigma_{nt}}. \quad (1)$$

The sample correlation-based score of the k -th ($k=1, \dots, K$) NWP model $\mathbf{f}_k(\cdot)$ is given by:

$$s_{nkt} = \mathbf{y}_{nt}^T \mathbf{f}_k(x_n, t), \quad (2)$$

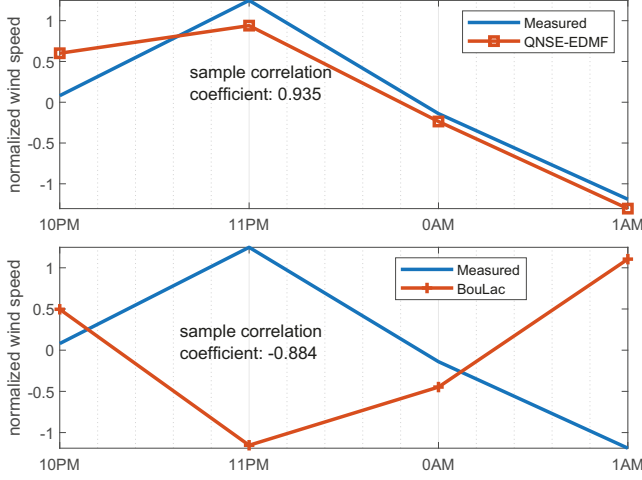


Fig. 5. Sample correlation coefficient-based score.

where $\mathbf{f}_k(x_n, t)$ is the centered and normalized wind speed by the k -th NWP model for the geographical location x_n (the latitude and longitude) of the grid cell that covers the n -th Mesonet site). It can be seen that $s_{nkt} \in [-1, 1]$, and a score of 1 indicates the k -th NWP matches perfectly with the n -th Mesonet measurements in its trend. Note that one NWP model would have totally N scores by the N Mesonet sites.

B. An Additive Ensemble Model

Generally, ensemble forecasting methods combine NWP models to produce a more accurate point forecast. One such approach is by model averaging [26], which may simply adopt equal weights for averaging or by following a certain criterion, e.g., Akaike's information criterion and Bayes' information criterion, for averaging. In the proposed approach, by leveraging the well-posed sample correlation coefficient-based scores, an additive ensemble model as a weighted combination of the NWP models is developed as follows:

$$\mathbf{F}_K(x) = \sum_{k=1}^K a_k \mathbf{f}_k(x), \quad (3)$$

where a_k is the weight for the k -th NWP model, x is the geographical index of the grid cell to be forecasted for. For brevity, the time index t is omitted from here on. Intuitively, the weight a_k should be higher when the scores of the k -th NWP model on all the N Mesonet sites are all high. Then, the objective boils down to finding the weights a_k for an optimal additive ensemble model $\mathbf{F}_K(x)$, which could be quantified by the following aggregate score by all the N Mesonet sites:

$$S_N = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_n^T \mathbf{F}_K(x). \quad (4)$$

With this insight, to find the weights a_k could be casted as an ensemble learning problem under the supervised learning framework, with the following aspects:

1) *Weak learner*: Each NWP model constitutes a weak learner. A weak learner produces wind speed forecasts for any location within the region. The weak learners are to be combined in an additive manner as in (3) towards obtaining a strong learner that produces more accurate forecasts.

2) *Meta supervisory data*: The wind speed measurements from the Mesonet sites comprise the meta supervisory data. Specifically, each individual Mesonet forms a 'training data point'. The wind speed measurements within the time window of T of a Mesonet site is used to score the weak learners. The score of a weak learner (an NWP model) on a training data point (a Mesonet site) is given by (2). It can be seen from (2) that the score of the weak learner takes values in $[-1, 1]$. A higher score indicates that the weak learner performs better on the training data point, in the sense that the forecast produced by the NWP model has a higher correlation with measured wind speed and thus correctly captures the trend for wind ramps. The scoring scheme is designed in analogy to the scores of binary classification models in supervised learning, in which a binary classification model scores 1 when its classification result matches the label of a training data point.

3) *Strong learner*: The strong learner is obtained from the additive model of the weak learners, as shown in (3). The additive coefficients, i.e., the weights of the weak learner, could be obtained through an ensemble-learning procedure using the training data points from Mesonet sites. Then, the strong learner as the additive model could be used to 'generalize' wind speed forecasts for any location (particularly for the wind farm locations) of the region. Further, the produced wind speed forecast is expected to be more accurate than the one produced by individual weak learners (NWPs).

C. An Ensemble Learning Method

The objective is to build an additive ensemble model from the K 'weak' NWP models, by using available N 'training data points' from the N Mesonet sites. Particularly, the score of the k -th 'weak' model on the n -th 'training data point' is given by s_{nk} in (2) which takes value in $[-1, 1]$. The weights a_k of each 'weak' model could be obtained by minimizing a well-defined cost function.

1) *A surrogate cost function*: based on the score function defined in (2), the following surrogate cost function parameterized by the weights $\mathbf{a}$ could be adopted:

$$C_N(\mathbf{F}_K; \mathbf{a}) = \sum_{n=1}^N d_n \log_2(1 + e^{-\mathbf{y}_n^T \mathbf{F}_K(x)}), \quad (5)$$

where $d_n = \frac{1}{N}$ is the data weight. The above cost function is adopted since $\log_2(1 + e^{-\mathbf{y}_n^T \mathbf{F}_K(x)})$ acting as an upper bound for $-\mathbf{y}_n^T \mathbf{F}_K(x)$. Therefore, minimizing $C_N(\cdot)$ forms a well-posed problem towards obtaining an additive ensemble model that has a high score on the 'training data points' of the Mesonet sites. Further, it is easy to see from (5) is convex, continuous, and Lipschitz differentiable w.r.t. to $\mathbf{F}_K$. Therefore, according to Theorem 1 of reference [27], $C_N(\cdot)$ could be efficiently minimized in a gradient descent (gradient boosting) manner.

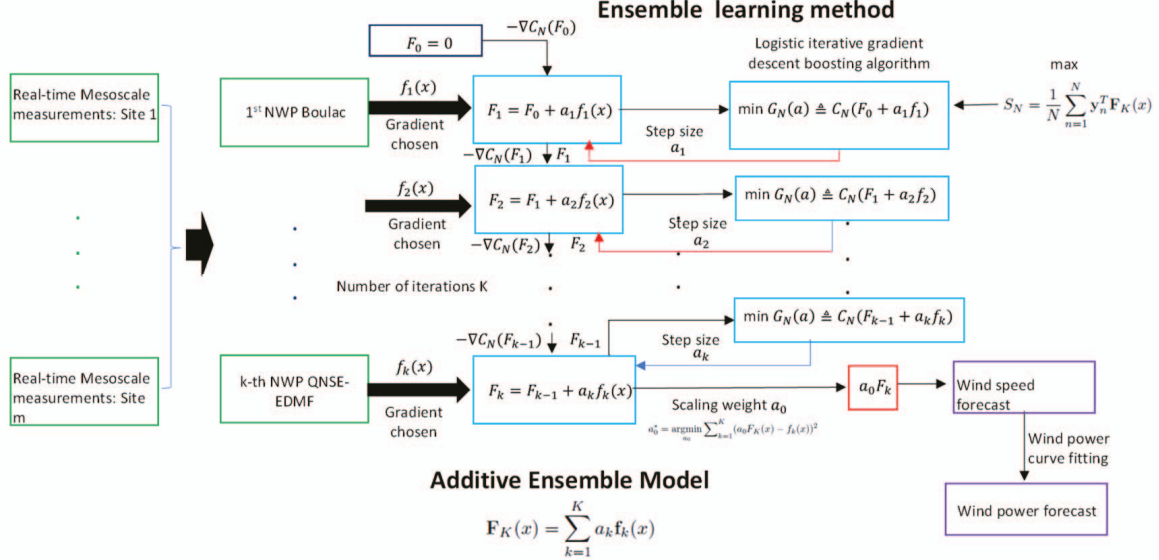


Fig. 6. Ensemble learning of NWP models.

2) *weights*: An iterative gradient-descent method similar to that in reference [28] could be adopted to solve for the weights a_k . Specifically, for the k -th step ($k=1, \dots, K$) of the iterations, an NWP model $\mathbf{f}_k$ is chosen so that it is closest to the negative gradient of C_N at $\mathbf{F}_{k-1}$, i.e., $-\nabla C_N(\mathbf{F}_{k-1})$. Further, the weight a_k is solved for, as the ‘step size’ for the corresponding ‘gradient’ $\mathbf{f}_k$, by minimizing $G_k(a) \triangleq C_N(\mathbf{F}_{k-1} + a \mathbf{f}_k)$. Then, the NWP model $\mathbf{f}_k$ is added to $\mathbf{F}_{k-1}$ to obtain the additive ensemble model of the $k+1$ -th step, i.e., $\mathbf{F}_k = \mathbf{F}_{k-1} + a_k \mathbf{f}_k$. More specifically, $\mathbf{F}_{k-1}$ for $k=1$ is set to zero. The closeness of $\mathbf{f}_k$ to $\nabla C_N(\mathbf{F}_{k-1})$ could be quantified by the following inner product between them:

$$\langle \mathbf{f}_k, -\nabla C_N(\mathbf{F}_{k-1}) \rangle = \frac{1}{\ln 2} \sum_{n=1}^N \frac{d_n s_{nk}}{1 + e^{-\mathbf{y}_n^T \mathbf{F}_{k-1}(x_n)}}, \quad (6)$$

Further, if $\langle \mathbf{f}_k, -\nabla C_N(\mathbf{F}_{k-1}) \rangle > 0$ holds, there exists a unique $a \in \mathcal{R}^+$ that minimizes $G_k(a)$, which is achieved at $G'_N(a) = 0$, which could be easily solved by using numerical methods. It is possible that $\langle \mathbf{f}_k, -\nabla C_N(\mathbf{F}_{k-1}) \rangle < 0$, i.e., even the best one among the remaining $K+1-k$ models contributes negatively to the present additive model, then, a_k would be negative, which can also numerically solved.

3) *Scaling weights*: The additive ensemble model in (3) is obtained by additively combining the weak learners according to their weights. It is noted that in classic binary classification problems, scaling the voting weights won’t change the classification result (as the classification decision is made by inspecting the sign of $\mathbf{F}_k$); however, the additive ensemble model in (3) in this work is used to produce wind speed forecasts, and thus the scales of the voting weights would have significant impact over the forecasting results. Therefore, an additional post-processing step is necessary to augment the output of the additive ensemble model. Specifically, a scaling factor a_0 is applied to the additive ensemble model, and it is

found by minimizing the mean squared error between all the individual NWP models, as follows:

$$a_0^* = \operatorname{argmin}_{a_0} \sum_{k=1}^K (a_0 \mathbf{F}_K(x) - \mathbf{f}_k(x))^2, \quad (7)$$

where x is the data of wind speed forecast within the time window of T for an individual wind farm.

4) *Training data weights*: For improved wind power forecasting performance, it could be necessary to train an additive ensemble model for each individual wind farm. One feasible method is to adapt the weights of the training data points, d_n , for each wind farm. For a wind farm, the wind speed measurements from the Mesonet sites in proximity would be of higher values than those from remote Mesonet sites. With this insight, distance-based data weights could be adopted:

$$d_n = \frac{1}{Z} \frac{1}{1 + e^{-\gamma \|\mathbf{x}_n - \mathbf{x}_f\|_2}}, \quad (8)$$

where $\mathbf{x}_n$ and $\mathbf{x}_f$ are the geographical coordinate (the latitude and longitude) of the n -th Mesonet site and a wind farm, respectively, Z is a normalizing constant such that $\sum_{n=1}^N d_n = 1$, and γ ($\gamma > 0$) is a control parameter that tunes the effect of the distance $\|\mathbf{x}_n - \mathbf{x}_f\|_2$ on the data weight d_n (particularly, when γ^{-1} is much smaller than distance, the data weights tend to be equal; and when γ^{-1} is much larger than distance, the data weights tend to be inversely proportional to distance). Similarly, correlation-based data weights could also be adopted by using the correlation of the wind speed measurements of Mesonet sites to wind power production of wind farms:

$$d_n = \frac{1}{Z} \frac{1}{1 + e^{-\gamma \mathbf{y}_n^T \mathbf{P}_f}}, \quad (9)$$

in which $\mathbf{y}_n$ is normalized wind speed measurements, $\mathbf{P}_f$ is wind farm power, and the parameters Z and γ are similar to that of the distance-based data weights.

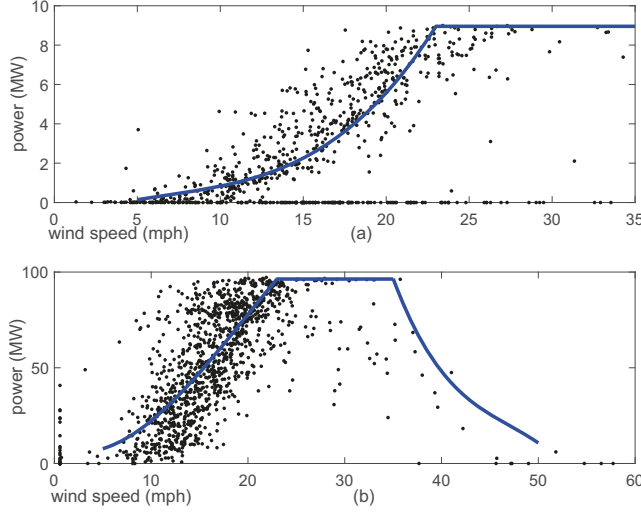


Fig. 7. Power curves for wind farm a) ‘BLSUMMIT’ and b) ‘ANACACHO’.

IV. NUMERICAL EXPERIMENT

A. Measurement and Experimental Data

Hourly wind power data for 135 wind farms in ERCOT within the covered region of the 86 Mesonet stations of year 2015, together with the corresponding hourly wind speed data of the Mesonet sites are used for numerical experiment. The hourly wind speed data is obtained by averaging over the 1-minute or 5-minute Mesonet measurements. In addition, re-analysis data for NWP is obtained by using eight physical schemes of WRF-ARW for an inner domain that covers the same geographical region with a grid cell size of 3km-by-3km and refreshing rate of 3 hrs. Using the latitude and longitude information, the grid cells that contain a Mesonet site or an ERCOT wind farm is identified, and its NWP data is retrieved and interpolated for the hub height of turbines.

Once the wind speed forecast is produced from the results of the NWP models, the wind power forecast for a wind farm is obtained by using the wind farm power curve. The wind farm power curve, which maps wind speed forecast to wind power forecast, is constructed by fitting a piecewise curve comprised of polynomial and linear segments to historical data on collated wind speed and power measurements in a least square manner. Two example power curves are illustrated in Fig. 7. It can be seen that the smaller wind farm ‘BLSUMMIT’ has a rated power output of 9 MW, and its power curve is closer to the manufacture’s power curve of a single wind turbine; while for the larger wind farm ‘ANACACHO’ rated at 99 MW, and the extended geographical area of the wind farm induces time lag between turbines reaching cut-out mode, leading to a smooth cut-out segment.

B. Test Results for A Single Wind Farm

The proposed methods are applied to the 2015 data of the wind farm ‘BRISCOE’ which contains 81 units and is rated at 150 MW. There were 1,523 instances of wind power ramp events that had an hourly change of 22.5 MW (i.e., 15% of

its rated capacity). Table I summarizes the test results. Two benchmark methods as the current practice or state-of-the-art are considered: 1) a single NWP model, and 2) a simple average of the ensemble (noted as ‘Ensemble-Ave’ in Table I). The single NWP model is chosen as the physical scheme ‘YSU’ which turned out to be the best among the ensemble in the test. Three proposed methods are also tested, which are with equal data weights, distance-based data weights, and correlation-based data weights (noted as ‘Proposed-ED’, ‘Proposed-DD’, and ‘Proposed-CD’ respectively in Table I). The control parameter γ is set to 0.735 which is the inverse of the standard deviation of the Mesonet coordinates for the method ‘Proposed-DD’, and to 1.863 which is the inverse of the mean correlation coefficient for the method ‘Proposed-ED’. Then, the normalizing parameter Z are calculated accordingly. The forecast error is measured in mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE). It can be seen that the ensemble methods significantly outperform the single best NWP model, and all the three proposed methods have comparable improvement over the state-of-the-art method ‘Ensemble-Ave’.

TABLE I
WIND POWER RAMP FORECAST ERROR FOR WIND FARM ‘BRISCOE’

	MAPE	MAE	RMSE
Single NWP	13.12 %	6.84 MW	8.96 MW
Ensemble-Ave	10.33 %	4.97 MW	6.33 MW
Proposed-DD	9.16 %	4.52 MW	5.35 MW
Proposed-ED	9.12 %	4.46 MW	5.16 MW
Proposed-CD	9.09 %	4.41 MW	5.07 MW

C. Test Results for A Region

Among the 135 wind farms, 97 wind farms that have correlation coefficients of 0.6 or above with at least four Mesonet sites in the year 2015 data are selected for testing the proposed methods. This selection is to ensure that the training data points formed by the Mesonet sites are sufficient and pertinent to the produced forecasting model for the selected wind farms. The proposed methods are tested on 852 large ramp events that has an hourly power ramp of 1,000 MW and above, including 453 up ramps and 399 down ramps. The results of forecast error in MAPE are summarized in Table. II with breakdown into up ramps and down ramps. Further, it is noted that the proposed methods utilize the real-time weather measurements in a time window to calculate the score according to (2). Thus, it would be interesting to investigate the impact of the duration of consecutive ramps on the accuracy of the proposed methods. The breakdown of wind power ramps by duration and the results on forecast error with regard to the ramp duration are illustrated in Fig. 8. For brevity, the up ramp and down ramp events are plotted in the same figure. Specifically, in Fig. 8, a duration of ‘-5’ indicates that a large down ramp occurs after four consecutive hours of wind power reduction. It can be seen that the large wind power ramp events are concentrated in the range of duration of 2-5 hours. Further, the proposed methods perform slightly better for wind power ramps with longer duration. This is because

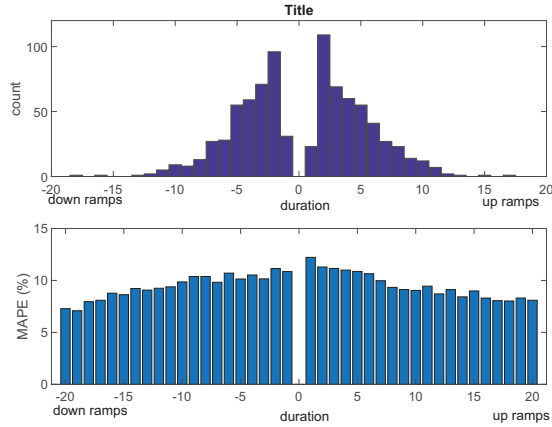


Fig. 8. (a) histogram of large wind power ramp by ramp duration (- for down ramps and + for up ramps) and (b) forecast error of Proposed-CD w.r.t. ramp duration.

the longer the wind ramp has been, the more trustworthy the correlation-based score is and thus the more accurate wind power ramp forecast is.

TABLE II
REGIONAL WIND POWER RAMP FORECAST ERROR (IN MAPE)

	Single NWP	Ensemble-Ave	Proposed-DD	Proposed-ED	Proposed-CD
overall	15.26%	12.61%	10.84%	10.95%	10.65%
up ramps	15.34%	12.64%	10.66%	10.78%	10.72%
down ramps	15.17 %	12.58 %	11.04 %	11.14%	10.57%

V. CONCLUSION

Meso-scale weather measurements contain pertinent information regarding the change of wind speed and wind power. Despite that NWP may have already incorporated these data into their input, the relatively low refreshing rate makes the most recent measurements not used by these models in a timely manner. In this study, an innovated method for weighted averaging of ensemble weather predictions according to their scores evaluated from the correlation with real-time measurements is developed. This method is non-intrusive, in the sense that it does not modify the output of individual NWP, which has great potentials to be adopted in applications. It is worth noting that although the presented numerical results are focused on hourly wind power ramp forecast, the proposed method can be used for wind power ramp forecasting at higher time resolutions. This is because the Mesonet real-time measurements are at 1-minute or 5-minutes timescale. Further, the ever-lasting expansion of the Mesonet system would enable wind power ramp forecast for more extended regions.

REFERENCES

[1] K. D. Orwig, et al., "Recent trends in variable generation forecasting and its value to the power system," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 924–933, July 2015.

[2] P. Lynch, "The origins of computer weather prediction and climate modeling," *Journal of Computational Physics*, vol. 227, no. 7, pp. 3431 – 3444, 2008.

[3] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1 – 8, 2012.

[4] H. Hughes, "Model output statistics forecast guidance," United States Air Force Environmental Technical Applications Center.

[5] P. Lynch, *Weather Prediction by Numerical Process*, ser. The Emergence of Numerical Weather Prediction. Cambridge University Press, 2006, pp. 1–27.

[6] E. Xydias, M. Qadrdan, C. Marmaras, L. Cipcigan, N. Jenkins, and H. Ameli, "Probabilistic wind power forecasting and its application in the scheduling of gas-fired generators," *Applied Energy*, vol. 192, pp. 382 – 394, 2017.

[7] D. J. Stensrud, *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2007.

[8] F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis, "The ECMWF Ensemble Prediction System: Methodology and validation," *Quarterly Journal of the Royal Meteorological Society*, vol. 122, no. 529, pp. 73–119, 1996.

[9] National Weather Service, "Forecast models," [Online] Available: <https://www.weather.gov/hnx/models>.

[10] —, "Mesoscale forecast models," [Online] Available: <https://www.weather.gov/mlb/localmesoscalemodels>.

[11] W. J. Shaw, L. K. Berg, J. Cline, C. Draxl, I. Djalalova, E. P. Grimit, J. K. Lundquist, M. Marquis, J. McCaa, J. B. Olson, C. Sivaraman, J. Sharp, and J. M. Wilczak, "The Second Wind Forecast Improvement Project (WFIP2): General Overview," *Bulletin of the American Meteorological Society*, Apr 2019.

[12] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state of the art in short-term prediction of wind power a literature overview, 2nd edition," [online] Available: ecolo.org/documents/documents_in_english/wind-predict-ANEMOS.pdf, Jan 2011.

[13] G. Giebel, J. Badger, L. Landberg, H. Nielsen, T. Nielsen, H. Madsen, K. Sattler, H. Feddersen, H. Vedel, J. Tfting, L. Kruse, and L. Voulund, "Wind power prediction using ensembles," Riso National Laboratory, Tech. Rep., Sept 2005.

[14] J. Taylor, P. Mcsharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *Energy Conversion, IEEE Transactions on*, vol. 24, pp. 775 – 782, Oct 2009.

[15] T. Heppelmann, A. Steiner, and S. Vogt, "Application of numerical weather prediction in wind power forecasting: Assessment of the diurnal cycle," *Meteorologische Zeitschrift*, vol. 26, no. 3, pp. 319–331, June 2017.

[16] M. Dabernig, G. J. Mayr, and J. W. Messner, "Predicting wind power with reforecasts," *Weather and Forecasting*, vol. 30, no. 6, pp. 1655–1662, 2015.

[17] G. Qu, J. Mei, and D. He, "Short-term wind power forecasting based on numerical weather prediction adjustment," in *2013 11th IEEE International Conference on Industrial Informatics (INDIN)*, July 2013, pp. 453–457.

[18] Q. Xu, D. He, N. Zhang, C. Kang, Q. Xia, J. Bai, and J. Huang, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1283–1291, Oct 2015.

[19] K. Higashiyama, Y. Fujimoto, and Y. Hayashi, "Feature Extraction of NWP Data for Wind Power Forecasting Using 3D-Convolutional Neural Networks," *Energy Procedia*, vol. 155, pp. 350 – 358, 2018.

[20] ERCOT, "Ercot using new forecasting tool to prepare for wind variability," [online] Available: www.ercot.com/news/releases/show/326, Mar 2010.

[21] AWS Truepower, "Description of the mesomap system," [online] Available: aws-dewi.ul.com/assets/Description-of-the-MesoMap-System1.pdf, Apr 2012.

[22] ERCOT, "Grid Information - Generation," [online] Available: www.ercot.com/content/cdr/html/CURRENT_DAYSTWPF.html.

[23] West Texas Mesonet, [online] Available: www.mesonet.ttu.edu/obsframe.html.

[24] Texas Tech Real Time Weather Prediction System, [online] Available: www.atmo.ttu.edu/bancell/real_time_WRF/TTUWRF-about.html.

[25] WRF-ARW: Advanced Research WRF model, [online] Available: www2.mmm.ucar.edu/wrf/users/.

- [26] C. G. H. Diks and J. A. Vrugt, "Comparison of point forecast accuracy of model averaging methods in hydrologic applications," *Stochastic Environmental Research and Risk Assessment*, vol. 24, no. 6, pp. 809–820, Aug 2010.
- [27] L. Mason, J. Baxter, P. Bartlett, and M. Frean, "Boosting algorithms as gradient descent," *Neural Inf. Process. Syst.*, pp. 512–518, 1999.
- [28] M. He, J. Zhang, and V. Vittal, "Robust online dynamic security assessment using adaptive ensemble decision-tree learning," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4089–4098, Nov 2013.