



Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales



Jie Zhang ^{a,*}, Mingjian Cui ^a, Bri-Mathias Hodge ^b, Anthony Florita ^b, Jeffrey Freedman ^c

^a University of Texas at Dallas, Richardson, TX, USA

^b National Renewable Energy Laboratory, Golden, CO, USA

^c State University of New York at Albany, Albany, NY, USA

ARTICLE INFO

Article history:

Received 29 June 2016

Received in revised form

28 November 2016

Accepted 20 January 2017

Available online 24 January 2017

Keywords:

Wind forecasting

Grid integration

Ramp forecasting

ERCOT

Optimized swinging door algorithm

ABSTRACT

The large variability and uncertainty in wind power generation present a concern to power system operators, especially given the increasing amounts of wind power being integrated into the electric power system. Large ramps, one of the biggest concerns, can significantly influence system economics and reliability. The Wind Forecast Improvement Project (WFIP) was to improve the accuracy of forecasts and to evaluate the economic benefits of these improvements to grid operators. This paper evaluates the ramp forecasting accuracy gained by improving the performance of short-term wind power forecasting. This study focuses on the WFIP southern study region, which encompasses most of the Electric Reliability Council of Texas (ERCOT) territory, to compare the experimental WFIP forecasts to the existing short-term wind power forecasts (used at ERCOT) at multiple spatial and temporal scales. The study employs four significant wind power ramping definitions according to the power change magnitude, direction, and duration. The optimized swinging door algorithm is adopted to extract ramp events from actual and forecasted wind power time series. The results show that the experimental WFIP forecasts improve the accuracy of the wind power ramp forecasting. This improvement can result in substantial costs savings and power system reliability enhancements.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

As an important renewable energy resource, wind power has been increasing dramatically in the electric power system. Currently, wind power meets approximately 4% of U.S. electricity demand [1]. However, some systems have noticeably higher amounts of wind installed; for example, the Electric Reliability Council of Texas (ERCOT) has experienced instantaneous wind power penetrations up to 29% [2]. Xcel Energy's Colorado system reached 60.5% hourly penetration [3]. The characteristics of upper availability limit, large variability, and uncertainty in wind power present a primary concern to system operators. Especially given the increasing amounts of wind power being integrated into the electric power system, accurate wind power forecasts are critically important for reliable and economic power system operations [4–6]. Although the current power grid is capable of handling small amounts of uncertainty and variability, ramp (or extreme)

events—such as sudden and large changes in wind power—are a critical concern for the system with high wind penetration. *Improving the accuracy of wind power forecasting is expected to reduce the discrepancy between the forecasted and actual wind power output, and thereby enhance the performance of wind power ramp forecasting and reduce wind integration costs.*

1.1. Wind power ramp forecasting

Large fluctuation incidents with large magnitudes and short durations, so-called ramping events, are a significant concern of power system operators. One of the biggest concerns associated with integrating a large amount of wind power into the grid is the ability to forecast and handle large ramps in wind power output. ERCOT has been experiencing a rapid growth of installed wind generation capacity and this tendency is expected to continue growing into 2017 [7]. A large down-ramp event occurred in the ERCOT system on February 26, 2008, which caused a system emergency [8]. Different time and geographic scales influence wind ramps, and both up and down ramps can have varying levels of severity. A similar challenge is also encountered in other power

* Corresponding author.

E-mail address: jiezhang@utdallas.edu (J. Zhang).

systems experiencing the rapid growth of renewable resources, e.g., California (Duck Curve [9]) and Hawaii. As the recent study conducted by the National Renewable Energy Laboratory (NREL) explored the implications and challenges of very high renewable electricity generation levels—from 30% up to 90%, focusing on 80%, of all U.S. electricity generation—in 2050, one great obstacle needed to solve for such a high penetration of renewables is to predict the ramping events accurately and understand their implications on the power system operations [10].

Different time and geographic scales influence wind ramps. Up and down ramps can have varying levels of severity, and generally, a down ramp could be more risky than an up ramp because of the availability of reserves. There are generally two main ways in which an inaccurate forecasting of ramp events can lead to large errors: *ramp magnitude* and *timing errors*. A magnitude error is defined as an event that is forecasted to occur at an expected time but with significantly different magnitude. In a ramp timing error, the actual ramp in power significantly leads/lags the forecasted ramp time. A correct forecast of large wind and solar ramps, especially about when they will happen, is imperative to reduce the wind integration costs and to ensure the reliability and security at independent system operators (ISOs). For example, an operator could decide to commit fewer MWs of fast-start units due to confidence on the near-term wind ramping capability forecasts, ultimately resulting in reduced system operation costs and increased system reliability.

Ramp forecasting is a relatively recent topic motivated by the need of improving the management of large and fast wind and solar power output variations, especially in a context of power systems with high renewable penetration. Thus, what constitutes the optimal output of a ramp-oriented forecasting tool is still not well defined. Generally, ramping events are parameterized by the following properties: *ramping start/end*, *ramping duration*, *ramping rate*, and *ramping magnitude*. Ramp event alerts, probabilistic ramp event occurrence, and ramp rate forecast represent some examples of the different ramp forecasting standpoints [11]. Ferreria et al. [12] provided a review of different ramp definitions and approaches to ramp event forecasting. Greaves et al. [13] defined a ramp as a change in wind power output that is at least 50% of the installed wind capacity and occurs within a time span of 4 h or less. Zheng and Kusiak [14] employed the rate of change of wind power output during a 10-min interval to define a ramp. Potter et al. [15] defined a ramp event as the change in power between two consecutive hours that is greater than or equal to 10% of the installed wind capacity. In a report by AWS Truewind [16], the up and down ramps were differently defined considering the different levels of risk: (i) a down ramp occurs if the power changes at least 15% of total capacity within 1 h; and (ii) an up ramp occurs if the power changes at least 20% of total capacity within 1 h. AWS Truewind also analyzed the ERCOT system, which defines a ramp event as a change of 20% or more of the rated capacity in any 30-min period [17]. It is important to note that a variety of ramp definitions in terms of capacity and temporal differences are commonly used in power systems.

Ramp forecasting can be characterized into two categories: indirect and direct forecasts. For *indirect ramp forecasting*, wind power forecasts can be produced by multiple forecasting methods such as numerical weather prediction (NWP), statistical, and machine learning methods. Ramp detection methods could be adopted in a post-processing method to detect ramping events in the forecasted wind and solar power. Both deterministic and probabilistic ramp forecasts can be produced in this way. For *direct ramp forecasting*, ramp extraction methods can directly be applied to historical measured and forecasted wind power to extract all historical ramping events. Statistical and machine learning methods can be developed based on the historical ramping events and ramp

forecasting errors to directly forecast ramping features (e.g., ramping magnitude, duration, and rate) at different timescales. This paper focuses on the indirect ramp forecasting method, with the aim to identify the benefits of improved wind power forecasting on extreme events.

Wind forecasting models are commonly divided into two categories based on the data utilized [18]. Statistical forecasting is based on the analysis of historical time series of wind; whereas physics-based forecasting uses numerical weather prediction (NWP) models that may include statistical corrections. The first type of forecasting model generally provides reasonable results in the estimation of long-term horizons, such as mean monthly, quarterly, and annual wind speed, or for very short timescales, such as a few hours [19]. The impact of atmospheric dynamics becomes more important for short-term horizons of a few hours to day ahead, and NWP models often produce more accurate forecasts on these timescales. Wind forecasts can have different forecasting objectives for multiple customers, including wind power plant owners, utility companies, and ISOs [20]. For ISOs, the capability to forecast rapid changes in wind power generation is a major concern; it is also the focus of this paper. This paper investigates the impacts of general wind power forecasting improvements on wind power ramp forecasting performance. A ramp extraction methodology is developed to compare the ramp forecasting accuracy of the improved wind power forecasts to the baseline forecasts.

1.2. Research motivation and objectives

The Wind Forecast Improvement Project (WFIP) was performed to improve short-term wind power forecasts and determine the value of these improvements to grid operators [21]. Large ramps can significantly influence system economics and reliability, upon which power system operators place primary emphasis. This paper evaluates the performance of wind ramp forecasting based on improved short-term wind power forecasts. The analysis is performed for the WFIP southern study region (Fig. 1) by comparing the experimental WFIP forecasts to the current short-term wind power forecasts (STWPF) at multiple spatial and temporal scales. The WFIP study region in ERCOT includes 8,296 MW of wind capacity (or approximately 85% of the total rated wind output) spread throughout 84 wind power plants. Short-term forecasts of 1- to 6-h-ahead (1HA to 6HA) wind power output are analyzed.

Following topics are discussed in the remainder of the paper: (i) the WFIP southern study region and the developed improved wind

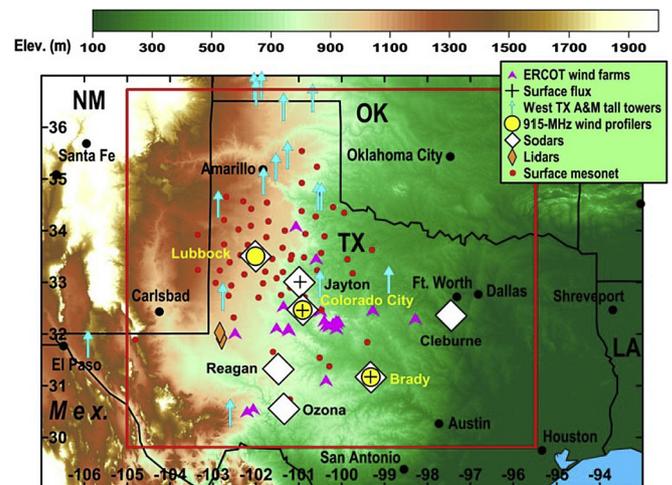


Fig. 1. WFIP Southern study region in ERCOT [22].

forecasting system; (ii) a statistical analysis to compare the performance of the WFIP to the current STWPF forecasts for ERCOT; (iii) the ramp extraction and forecasting methodology and the metrics for assessing ramp forecasting; and (iv) results and discussion of the ramp forecasting as a result of improved wind power forecasts.

2. Wind Forecast Improvement Project (WFIP)

WFIP encompassed two study regions: the northern study region and the southern study region [21,22]. In this paper, we focus on the ramp forecasting performance in the WFIP southern study region.

2.1. WFIP southern study region

ERCOT manages the flow of electric power to 24 million Texas customers - representing 85% of the state's electric load. The WFIP southern study region covers most of the ERCOT service area, as shown in Fig. 1. As of May 2015, ERCOT had more than 12,000 MW of wind capacity installed and a 40.58% wind penetration record happened on March 29, 2015 [2]. In this study, 1HA to 6HA wind power forecasts were generated by using both the WFIP and STWPF systems for a nearly 12-month period from October 2011 to mid-September 2012.

2.2. Wind forecasting system

The existing ERCOT's wind power forecasting system, short-term wind power forecast (STWPF), is used as a baseline for this study [22]. The STWPF used Mesoscale Atmospheric Simulations System model forecasts with initial conditions and boundary conditions from the Global Forecast System and the North American Mesoscale Model. Ensemble methods have been shown to produce more accurate forecasts. The WFIP experimental forecasting system consists of an ensemble of high-resolution rapid-update NWP models. Each of these ensemble members incorporates a variety of model configurations, physics parameterizations, and data assimilation techniques. The purpose of integrating all of these ensemble members into one system is to construct an optimized composite forecast, being able to predict forecast uncertainty and assess the relative performance of different modeling approaches. Fig. 2

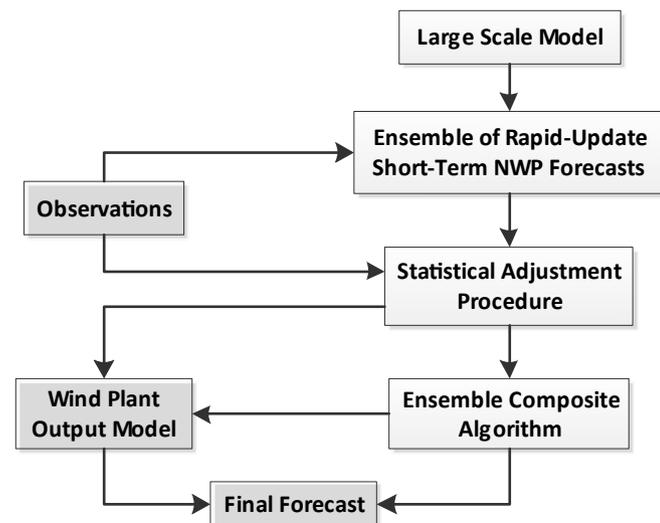


Fig. 2. Overall framework of the wind power forecasting system [22].

shows the overall framework of the wind power forecasting system. The WFIP ensemble members include [22]:

- i. The National Oceanic and Atmospheric Administration's 3-km High-Resolution Rapid Refresh model updated hourly;
- ii. Nine NWP models updated every 2 h on a 5-km grid:
 - (a) Three configurations of the Advanced Regional Prediction System [23,24];
 - (b) Three configurations of the Weather Research and Forecasting model [25];
 - (c) Three configurations of the Mesoscale Atmospheric Simulations System [26]; and
- iii. An Advanced Regional Prediction System model updated every 6 h on a 2-km grid.

The data from additional sensors deployed for this project, as well as tower data from a set of participating wind power plants within Texas, were assimilated into most of the ensemble members; however, the data from the project sensors were withheld from some ensemble members to gauge their impact on the forecasts [22].

A model output statistics (MOS) procedure was applied to the forecasts from each NWP system. The MOS is designed to correct systematic errors of relevant NWP meteorological variables (e.g., wind speed and direction) at forecast sites. In the WFIP system, a screening multiple linear regression approach was used for each wind power plant site and each NWP model. The purpose of MOS is to remove systematic errors due to unresolved sub-grid processes, limitations in model physics, or data assimilation techniques. There is a lag before NWP forecast data are available due to the time required for data gathering, analysis, initialization, and model execution. Thus, there is typically more recent data available than what is used to initialize the latest available NWP run. There are, however, time series prediction schemes that take advantage of the newer data to improve the NWP forecast performance for the 0- to 2-h look-ahead period. In addition to the MOS procedure, the ensemble includes statistical predictions based purely on recent time series of observational data, known as the persistence adjust (PA) method [18]. The PA method determines the initial forecast bias at the time of forecast generation and then applies a static bias correction to the remainder of the 6-h forecast. The statistical adjustment procedure corrects the initial bias and applies a separate bias correction independently to each forecast interval.

The MOS output for the individual NWP systems was then used as input to an optimized ensemble model, which created a composite deterministic or probabilistic forecast from the set of MOS-adjusted NWP forecasts. For a deterministic forecast, the optimized ensemble model training strategy was based on a rolling sample of the last 30 days to weight each individual forecast according to its performance. The strategy is to assign large weights to forecast members that are likely to perform better based on previous forecasts and observations in the training sample. More details of the forecasting system can be found in Ref. [22].

2.3. Wind power forecasting performance comparison between WFIP and STWPF

A comprehensive statistical analysis was performed in Ref. [27] to evaluate the forecasting improvements provided by WFIP. Standard statistical metrics (i.e., Pearson's correlation coefficient, root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE)) and distributions of wind power forecasting errors were used to compare the performance of WFIP and STWPF at different forecast horizons. Results showed that the experimental WFIP provided better results than the current STWPF at all

forecasting timescales.

To extend the results from Ref. [27] and further illustrate the forecasting trend from 1HA to 6HA timescales, Fig. 4 shows the wind power forecasting error percentiles (from the 10th to 90th percentile) at different forecasting horizons. It is observed in Fig. 4 that the 10th to 90th percentiles of forecasting errors continuously decrease with the forecasting horizon. A comparison of the STWPF (in Fig. 3(a) and Fig. 4(a)) to the WFIP forecasts (in Fig. 3(b) and Fig. 4(b)) shows that the 1HA to 6HA forecasts from the WFIP performed significantly better than the 1HA to 6HA forecasts from the STWPF.

3. Ramp extraction and forecasting methodology

In this paper, we employed four definitions of significant wind power ramps based on the power change magnitude, direction, and duration. The optimized swinging door algorithm (OpSDA) [28,29] was adopted to extract the ramp events from both the actual and forecasted wind power series at different timescales. A suite of metrics was proposed or adopted to evaluate the performance of ramp extraction and forecasting.

Significant wind power ramps can be defined based on the power change magnitude, direction, and duration. The same four definitions proposed in Ref. [27] are investigated in this paper:

- Significant Ramp Definition 1 – the change in wind power output is greater than 30% of the installed wind capacity without constraining the ramping duration.
- Significant Ramp Definition 2 – the change in wind power output is greater than 25% of the installed wind capacity within a time span of 4 h or less.
- Significant Ramp Definition 3 – a significant ramp rate is defined as the change rate in wind power output that is greater than 10% of the installed wind capacity per hour.
- Significant Ramp Definition 4 – a significant up-ramp is defined as the change in wind power output greater than 20% of wind capacity within a time span of 4 h or less; a significant down-ramp is defined as the change in wind power output greater

than 15% of the installed wind power capacity within a time span of 4 h or less [16].

3.1. Ramp extraction using the optimized swinging door algorithm (OpSDA)

Ramps are extracted through a linear piecewise approximation to the original time series of actual or forecasted wind power in this study. The swinging door algorithm (SDA) was adopted previously in Ref. [27] to extract wind power ramps. However, the standard SDA did not optimally determine the tunable parameter and optimally segregate the wind power signal. To improve the accuracy of ramps extraction, in this paper we adopt a recently developed ramp extraction method, the optimized SDA (OpSDA) [28,29]. To determine any significant ramp or ramp rate as defined by definitions 1 to 4, the start and end points of all ramps in a given time series of wind power need to be identified. Toward this end, the OpSDA was adopted to extract ramp periods in a series of power signals by identifying the start and end points of each ramp. The OpSDA is capable of automatically addressing the changes in ramp definitions in terms of capacity and temporal differences.

The swinging door algorithm (SDA) allows for the consideration of a threshold parameter influencing the algorithm's sensitivity to ramp variations. The OpSDA improves the performance of the original SDA through optimization. First, the SDA is utilized to segregate the wind power generation into consecutive segments in a piecewise linear fashion. Then a dynamic programming approach is used to combine adjacent segments into significant ramps when the decision thresholds are met. An increasing length score function, S , is designed based on the length of the interval segregated by the SDA. The dynamic programming seeks to maximize the length score function, which corresponds to a ramp event. Fig. 5 illustrates the overall structure of the OpSDA to identify significant ramps. Based on the four ramp definitions, significant ramps can be identified. Each significant ramp is characterized by the (i) start and end hours; (ii) wind power at the start and end points; and (iii) direction of the ramp. An increasing length score function, S , is

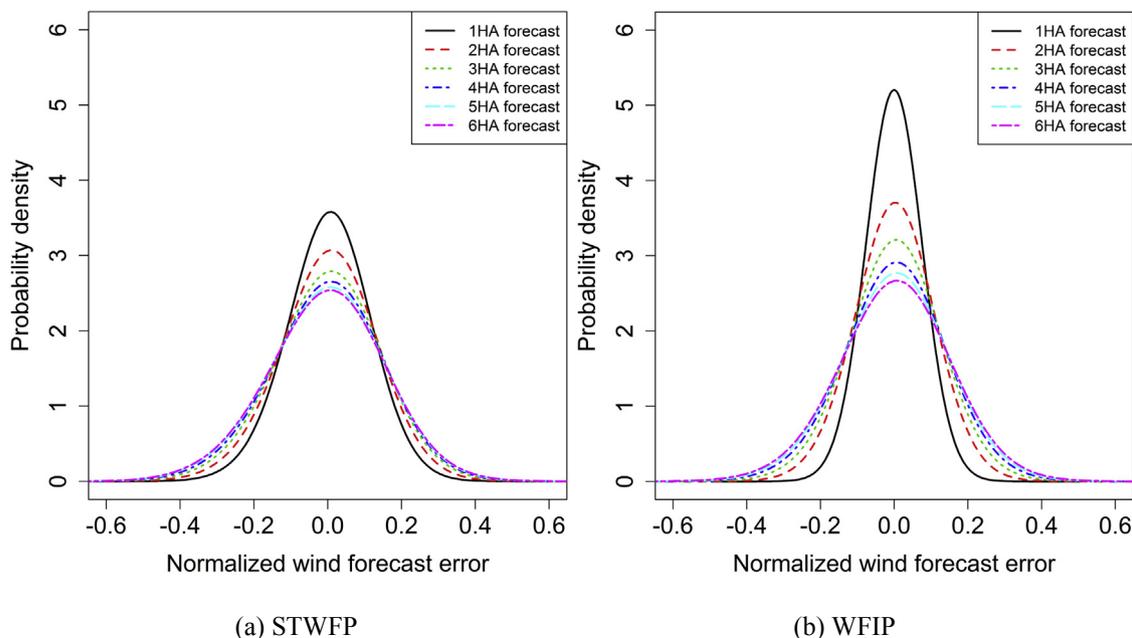


Fig. 3. Distributions of wind power forecasting errors at different forecasting horizons.

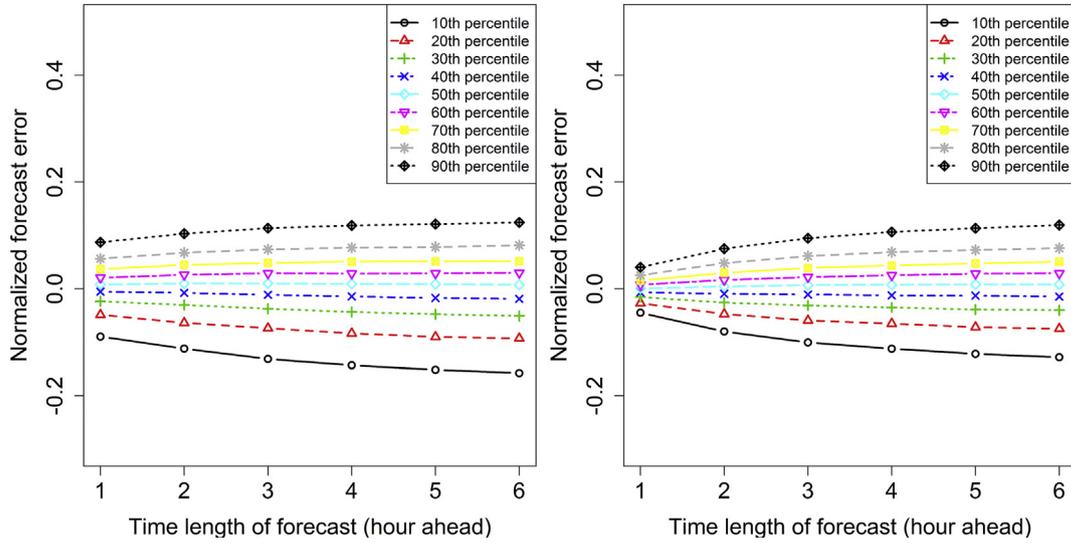


Fig. 4. 10th to 90th percentiles of wind power forecasts at different forecasting horizons.

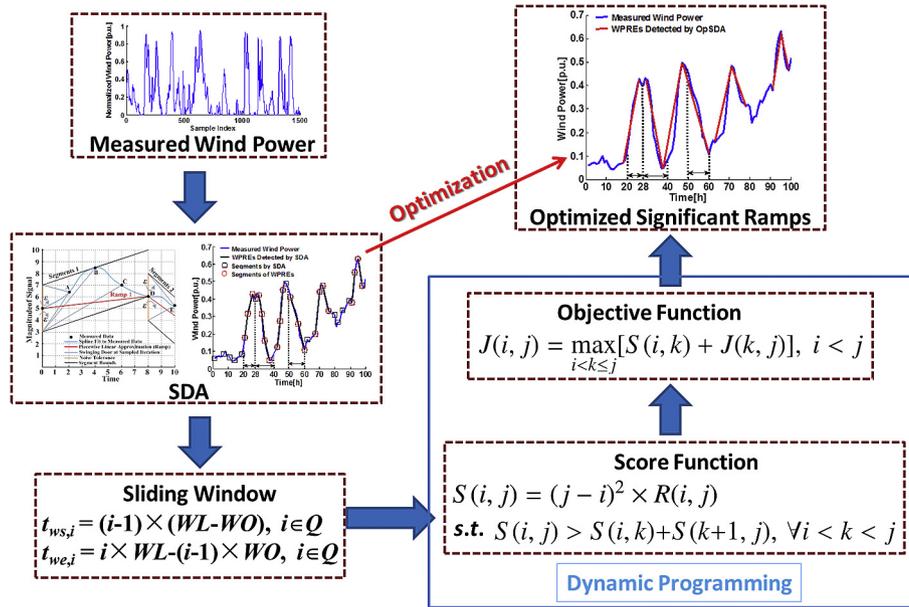


Fig. 5. The OpSDA for the extraction of wind power ramps.

designed based on the length of the interval segregated by the SDA. Given a time interval, (i, j) , of all discrete time points and an objective function, J , of the dynamic programming algorithm, a wind power ramping event is detected by maximizing the objective function:

$$J(i, j) = \max_{i < k < j} [S(i, k) + J(k, j)], \quad i < j \quad (1)$$

s.t.

$$S(i, j) > S(i, k) + S(k + 1, j), \quad \forall i < k < j \quad (2)$$

$$S(i, j) = (j - i)^2 \times R(i, j) \quad (3)$$

where $J(i, j)$ can be computed as the maximum over $(j - i)$ sub-problems. The term of $S(i, k)$ is a positive score value corresponding to the interval, (i, k) , which conforms to a super-additivity property

in Eq. (2). There is a family of score functions satisfying Eq. (2), and the score function expressed in Eq. (3) is adopted in this paper. $R(i, j)$ represents a ramp within the time interval (i, j) . A detailed description of the OpSDA can be found in Refs. [28,29].

3.2. Metrics for evaluating significant ramp forecasting performance

A suite of ramp event detection metrics is used to evaluate the performance of ramp forecasting, including a contingency table, categorical statistics, and performance diagrams. Categorical statistics provide measures of accuracy and skill for forecasts of notable events, such as ramps in power, detrimental temperatures, or rainfall. The adopted metrics include the probability of detection (POD), the critical success index (CSI), the frequency bias score (FBIAS), and the success ratio (SR) [12,27]. They are calculated based on a contingency table in Table 1 that provides a measure of

Table 1
Contingency table for ramp events observation and forecast.

	Observed YES	Observed NO	Total
Forecast YES	TP (hits)	FP (false alarm)	TP + FP
Forecast NO	FN (misses)	TN	FN + TN
Total	TP + FN	FP + TN	$N = TP + FP + FN + TN$

Table 2
Number of observed ramps in the actual wind power for all wind farms.

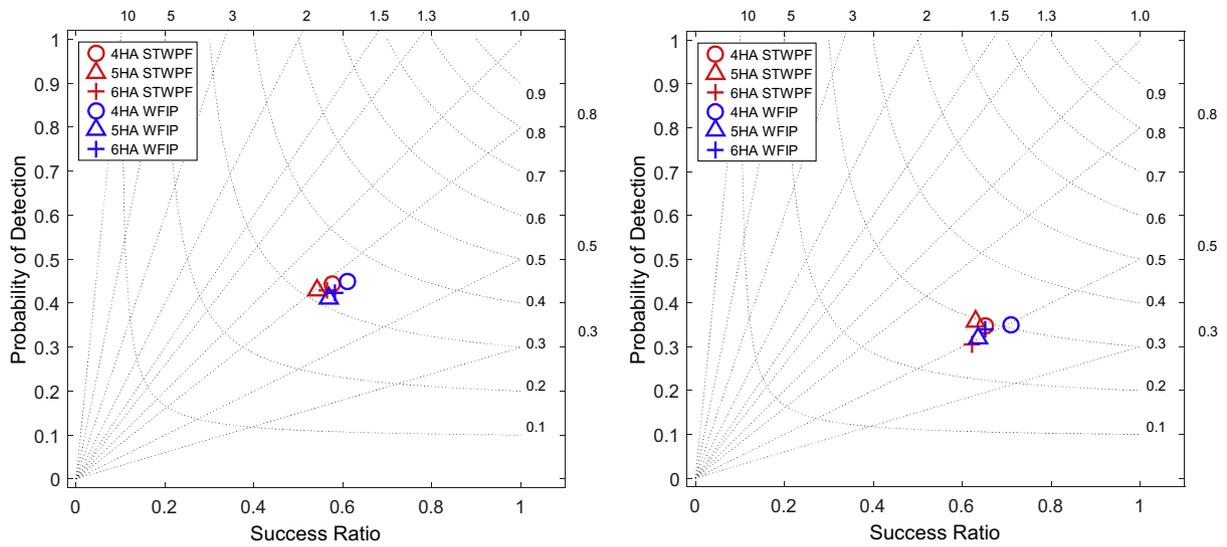
Ramp Type	Ramp Def. 1	Ramp Def. 2	Ramp Def. 3	Ramp Def. 4
Up Ramps	187	159	208	245
Down Ramps	160	126	150	335

skill for the forecasting ramps approaching the actual ramps. True positive (TP) represents the number of forecasted ramps (forecast YES) that are actually observed in the actual power output

(observed YES); false positive (FP) is the number of forecasted ramps that are not observed in the actual wind power (observed NO); false negative (FN) represents the number of observed ramps (observed YES) that are not predicted by the wind forecasting system (forecast NO); true negative (TN) is the number of non-occurring events for both observed and forecasting results; and N is the total number of events. The relationship among the POD, CSI, FBIAS, and FAR can be visualized on a *performance diagram* [30].

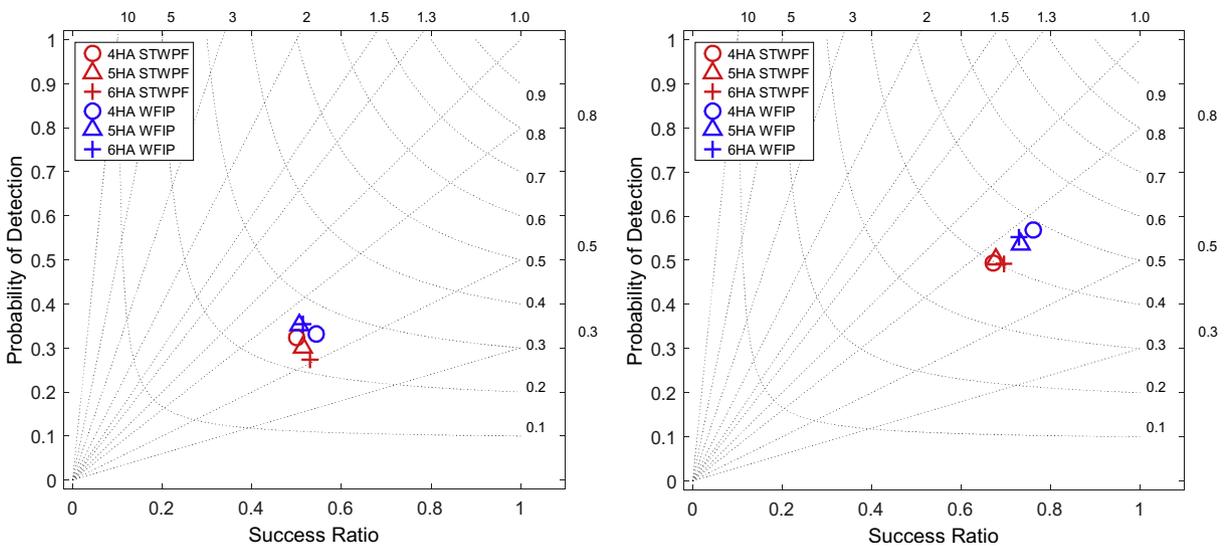
Categorical statistics provide measures of accuracy and skill for forecasts of notable events, such as ramps in power, detrimental temperatures, or rainfall. Based on the contingency table, a suite of metrics can be derived for ramp forecasting performance evaluation, given as follows.

Probability of detection (POD) is defined as the ratio between the number of true positives and the number of observed positives, which indicates the fraction of observed YES events that are actually forecasted.



(a) Ramp Definition 1: Ramp magnitude only

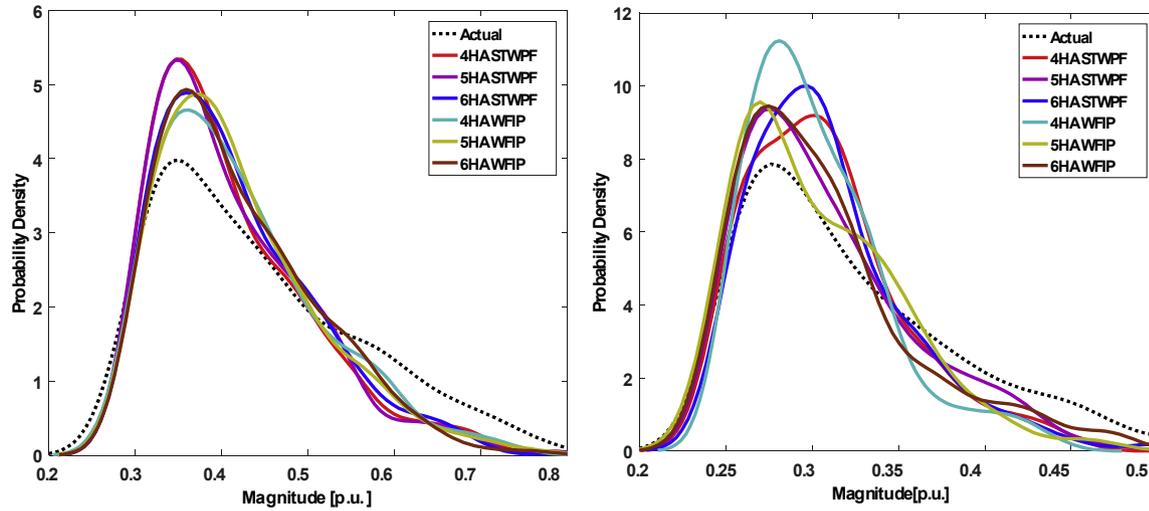
(b) Ramp Definition 2: Ramp magnitude and duration



(c) Ramp Definition 3: Ramp rate

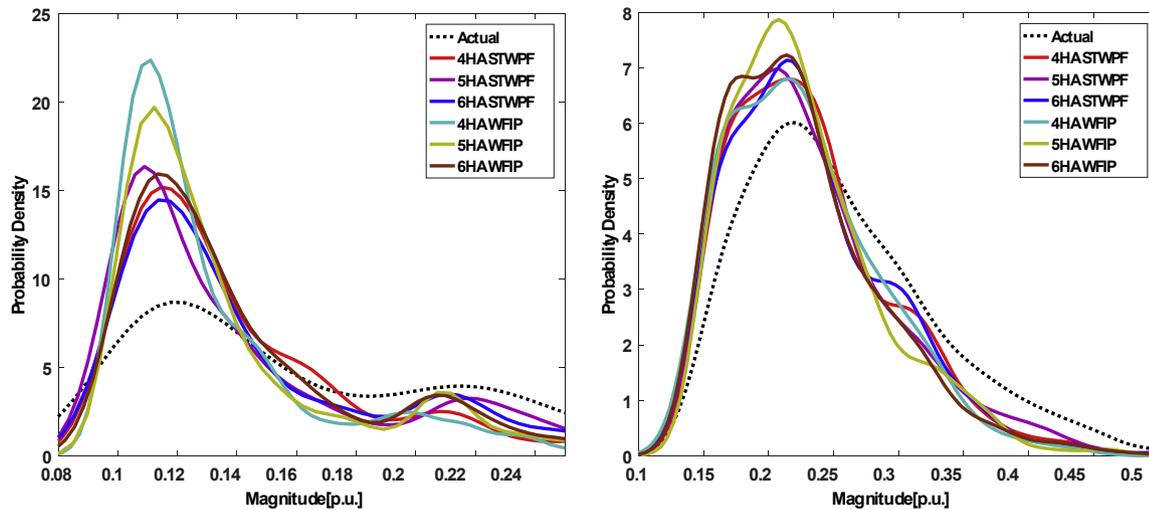
(d) Ramp Definition 4: Ramp direction, magnitude, and duration

Fig. 6. Comparison of annual ramp forecasting performance of STWPF to WFIP for different forecast timescales and ramp definitions.



(a) Ramp Definition 1: Ramp magnitude only

(b) Ramp Definition 2: Ramp magnitude and duration



(c) Ramp Definition 3: Ramp rate

(d) Ramp Definition 4: Ramp direction, magnitude, and duration

Fig. 7. Comparing annual distribution of ramp magnitude between STWPF and WFIP.

$$POD = \frac{TP}{TP + FN} \quad (4)$$

Critical success index (CSI) is used to measure the fraction of observed and/or forecasted events that are correctly predicted, given by

$$CSI = \frac{TP}{TP + FN + FP} \quad (5)$$

The value of CSI is between 0 and 1, with 1 representing a perfect prediction. The CSI considers only observed and forecasted ramps, excluding true negative events.

Frequency bias score (FBIAS) measures the ratio of the frequency of forecasted YES events to the frequency of observed YES events.

$$FBIAS = \frac{TP + FP}{TP + FN} \quad (6)$$

The ramp forecasting system tends to underforecast when

$FBIAS < 1$, and it tends to overforecast when $FBIAS > 1$.

False alarm ratio (FAR) measures the fraction of predicted YES events that did not occur, given by

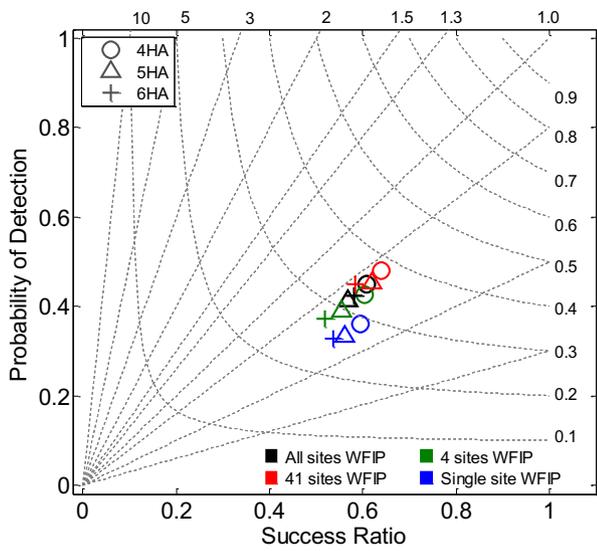
$$FAR = \frac{FP}{FP + TP} \quad (7)$$

The metric success ratio (SR) is calculated from FAR by subtracting it from 1. SR measures the fraction of predicted YES events that occurred.

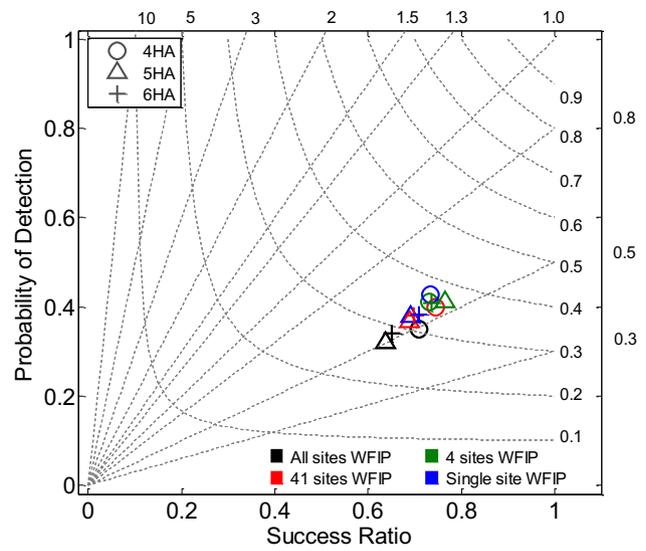
The relationship among the POD, CSI, FBIAS, and FAR can be visualized on a performance diagram based on

$$CSI = \frac{1}{\frac{1}{POD} + \frac{1}{1-FAR} - 1} \quad (8)$$

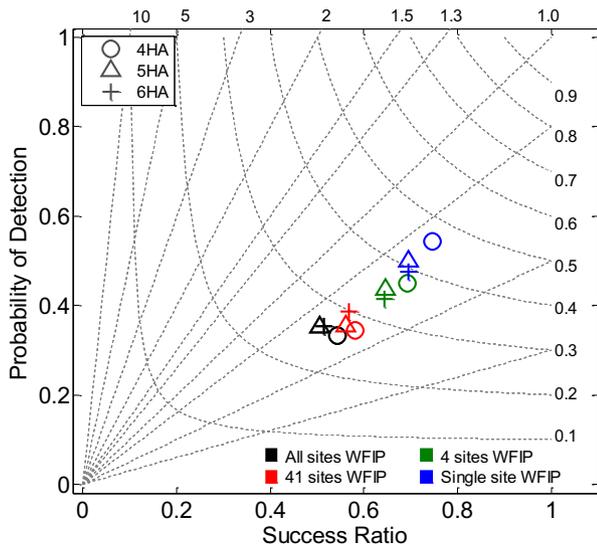
$$FBIAS = \frac{POD}{1 - FAR} \quad (9)$$



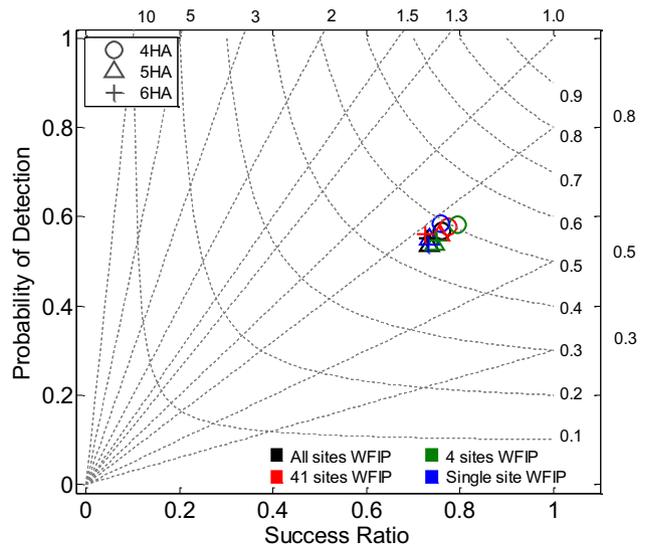
(a) Ramp Definition 1: Ramp magnitude only



(b) Ramp Definition 2: Ramp magnitude and duration



(c) Ramp Definition 3: Ramp rate



(d) Ramp Definition 4: Ramp direction, magnitude, and duration

Fig. 8. Comparison of WFIP ramp forecasting performance for different spatial aggregations.

Table 3
Number of ramps for different spatial aggregations with different ramp definitions.

Forecasting Timescale	Spatial Aggregation	Ramp Def. 1	Ramp Def. 2	Ramp Def. 3	Ramp Def. 4
4HA	1 site	635	1049	1105	853
	4 sites	514	854	634	684
	41 sites	441	689	305	545
	All sites	436	607	229	474
5HA	1 site	625	1007	1098	836
	4 sites	515	815	669	667
	41 sites	448	657	323	526
	All sites	427	584	267	450
6HA	1 site	615	992	1038	819
	4 sites	525	817	654	658
	41 sites	441	687	344	556
	All sites	418	596	266	474

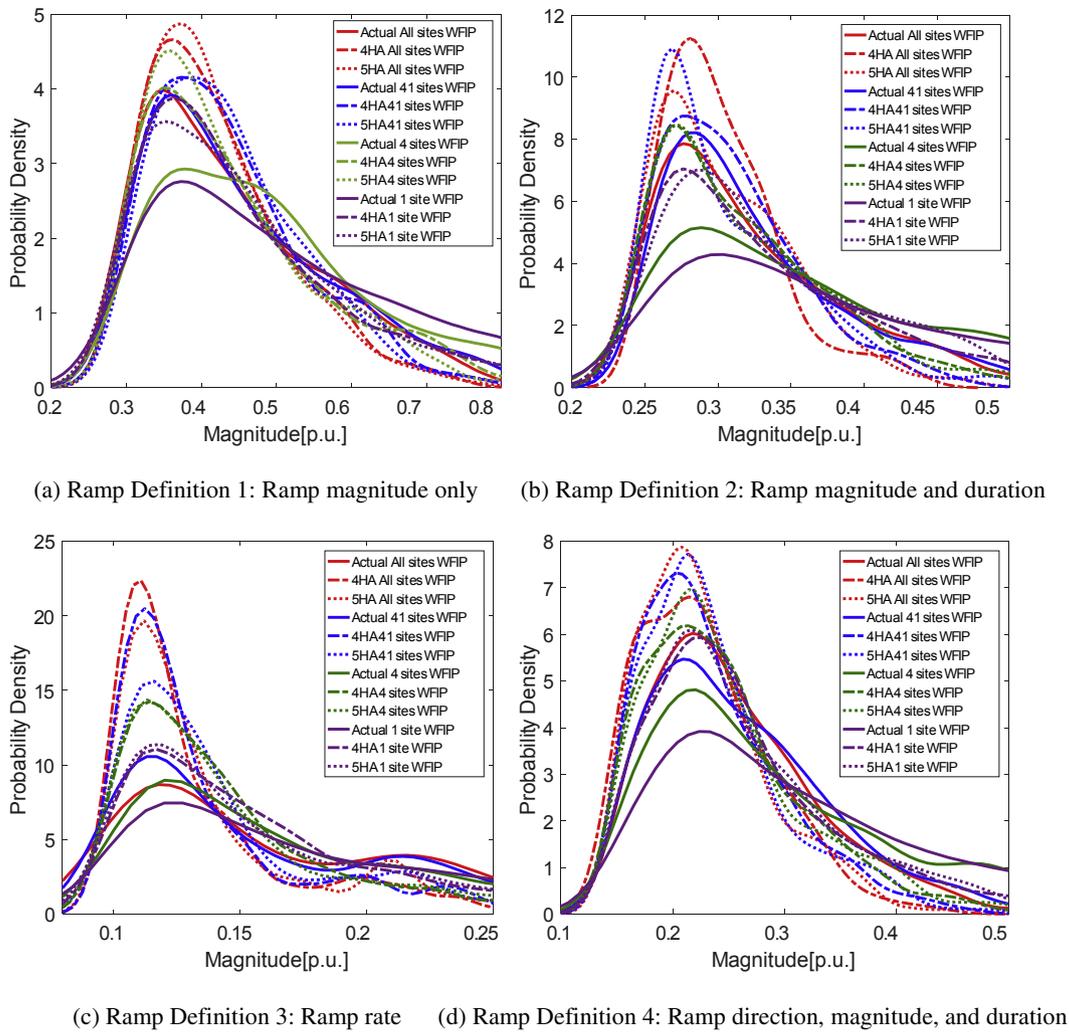


Fig. 9. Comparing WFIP distribution of ramp magnitude for different spatial aggregations.

4. Evaluating significant ramp forecasting performance based on improved wind power forecasts

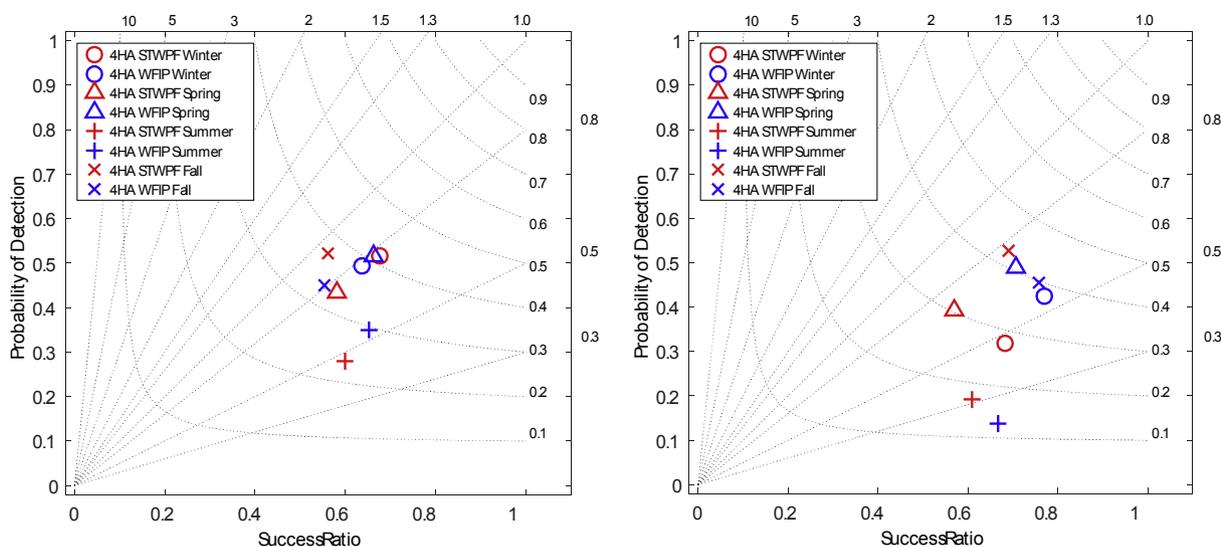
The performance of wind power ramp forecasting is affected by many factors such as the ramp extraction algorithm, wind power forecasts, wind penetration level in the power system, and others. In this paper, we focus on analyzing the effects of the wind power forecasting improvements on the performance of ramp forecasting. The true positive forecast defined by Greaves et al. [13] is adopted and modified in this study. Greaves et al. [13] defined a true positive forecast as a forecast ramp with a measured ramp of the same direction (either up or down) within ± 12 h of the time of the forecast ramp. In this study, a true positive forecast is defined as a forecast ramp with a measured ramp of the same direction within ± 6 h of the time of the forecast ramp. This ± 6 h range could provide sufficient data for temporal uncertainty analysis and maintain a realistic connection between forecast and measured significant wind ramp events. We have analyzed the ramp forecasting performance at different spatial and temporal scales. The annual, seasonal, and monthly ramp forecasting evaluation results are analyzed and discussed in the following sections. In this paper, the notation “annual” represents the nearly 12-month period from October 2011 to mid-September 2012.

4.1. Annual ramp forecasting performance

Table 2 lists the number of significant up and down ramps in actual wind power generation throughout the whole year for all wind farms. It is seen that there are generally more up ramps than down ramps according to the four significant ramp definitions. By comparing the four definitions, there are more ramps fitting to Ramp Definition 4 than to other definitions.

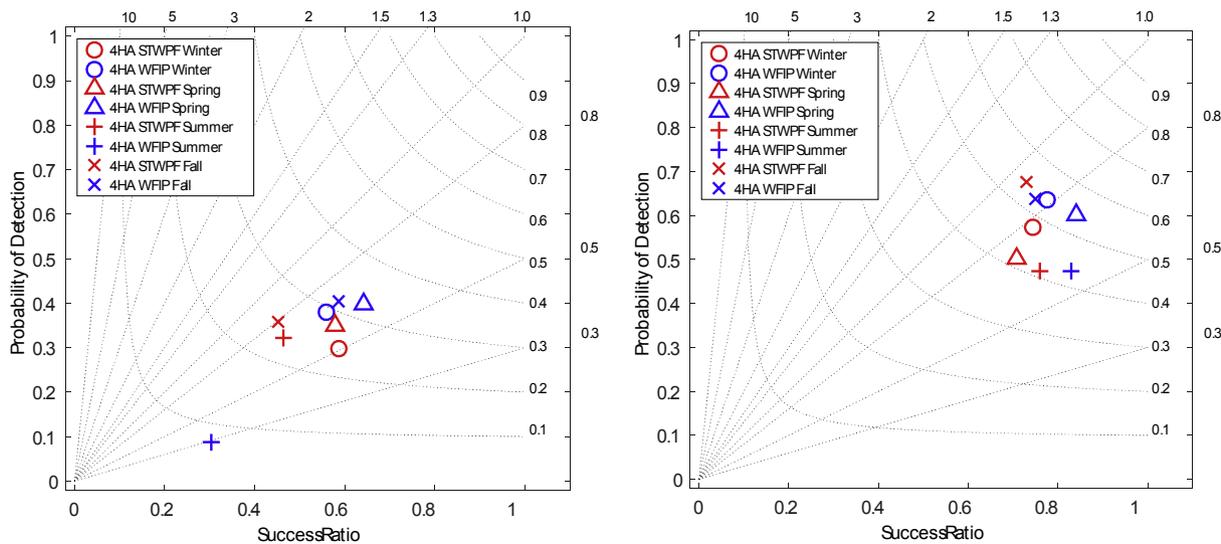
A performance diagram (as illustrated in Fig. 6) is used for quantitative and visual analysis, thereby understanding whether the wind power ramp forecasting performance is improved. Fig. 6 illustrates the annual ramp forecasting performance of STWPF and WFIP for different forecasting timescales and significant ramp definitions. In the performance diagrams, (i) the left axis represents the value of POD; (ii) the bottom axis represents the success ratio; (iii) the diagonal dashed lines represent FBIAS with the values shown on the right and top axes; and (iv) the solid curves show CSI with the values on the right-inside graph border. The figure shows the 4HA, 5HA, and 6HA forecasts. The performance diagram shows the ramp forecast performance space—that is, the overall ramp forecast metrics are improved as the forecast moves toward the upper right of the diagram.

It is observed in Fig. 6(a) according to Ramp Definition 1 that: (i) the 4HA WFIP has a larger success ratio, POD, and CSI values



(a) Ramp Definition 1: Ramp magnitude only

(b) Ramp Definition 2: Ramp magnitude and duration



(c) Ramp Definition 3: Ramp rate

(d) Ramp Definition 4: Ramp direction, magnitude, and duration

Fig. 10. Comparison of seasonal ramp forecasting performance of STWPF to WFIP for 4HA forecasts.

than the 4HA STWPF; (ii) the 5HA WFIP also has a larger success ratio, however a smaller POD, than the 5HA STWPF; and (iii) the 6HA WFIP has a larger success ratio and CSI values than the 6HA STWPF. Similar results are also observed according to the Ramp Definition 2 as shown in Fig. 6(b). With Ramp Definitions 3 and 4, WFIP performs better than STWPF at all forecast timescales. Overall, the ramp forecasting as a result of the improved wind power forecasts has been improved. The FBIAS values are smaller than one for all cases, which indicates that both the WFIP and STWPF ramp forecasts tend to underforecast. This underforecasting trend could alert the power system operators to employ appropriate strategies to compensate for wind power ramps, such as holding more reserves.

Fig. 7 shows the distribution of the ramp magnitude. For ramp Definitions 1 and 2, at the peak of the distribution, both WFIP and STWPF present larger ramp magnitude comparing to the actual wind power ramps. For Definition 3 in Fig. 7(c), the

ramp magnitude distribution presents two peaks in the actual ramps, which are also successfully forecasted by both WFIP and STWPF.

4.2. Ramp forecasting performance at multiple spatial scales

Understanding the ramp forecasting performance within different spatial scales can provide a better understanding of the flexibility requirements and reliability impacts of wind integration on the grid. Four scenarios are analyzed based on the number of wind power plants, including: (i) all wind power plants within the WFIP southern study region with a 8,296-MW capacity; (ii) 41 wind power plants with an aggregated 4,863-MW capacity; (iii) 4 wind power plants with an aggregated 1,073-MW capacity; and (iv) a single wind power plant with an approximate 500-MW capacity. Fig. 8 compares WFIP ramp forecasting performance at multiple spatial and temporal scales with different significant ramp

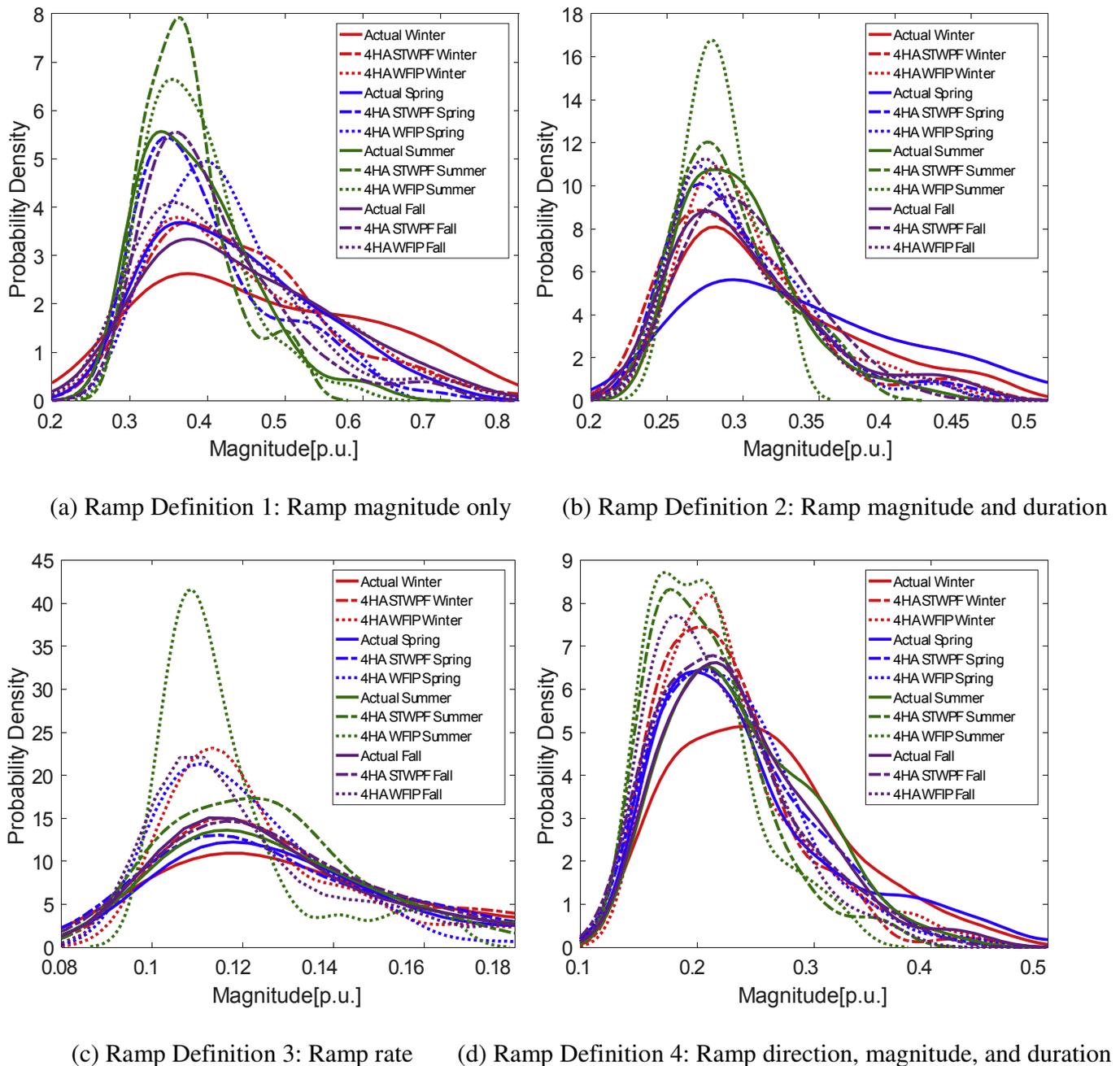
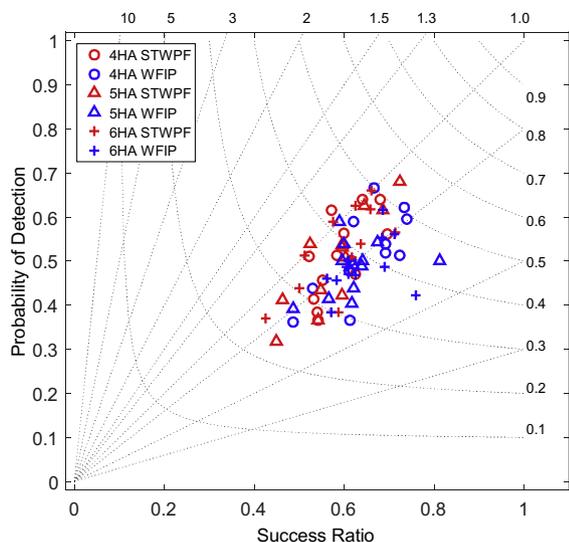


Fig. 11. Comparing distribution of ramp magnitude between STWPF and WFIP for 4HA forecasts.

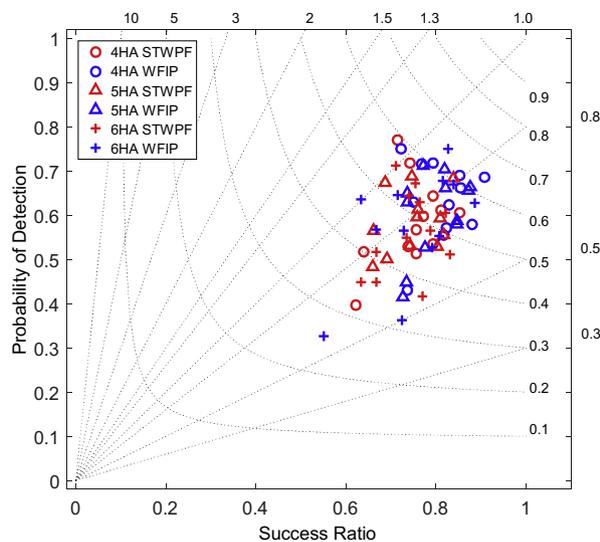
definitions. It is observed that, for Definition 3, the success ratio decreases with aggregating more wind power plants, indicating that there are more predicted YES events that actually occurred for a small wind power capacity. For Definitions 2 and 3, as shown in Fig. 8(b) and (c), respectively, the ramp forecasting for the single wind power plant also presents relatively larger POD values, indicating that there are more observed YES events that are actually forecasted. However, the ramp forecasting with more aggregated wind power plants (all plants and 41 plants scenarios) presents relatively larger POD and success ratio values for Definition 1, as shown in Fig. 8(a). It is shown that the performance of ramp forecasting is sensitive to the significant ramp definition. For power system operations, it is important to make multiple strategies to handle different types of wind power ramps based on the definition

of significant ramps. Table 3 lists the number of significant ramps for different spatial aggregations with different significant ramp definitions. It is seen that the number of significant ramps is decreasing by aggregating more wind power plants, which shows the smoothing effect from geographic diversity in wind power ramps.

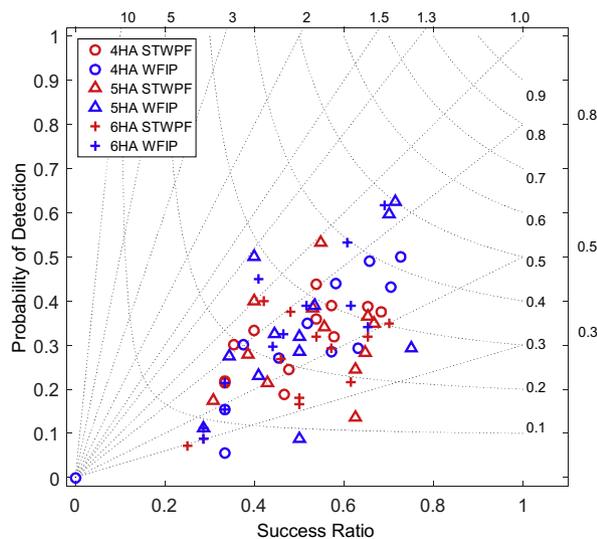
Fig. 9 shows the distribution of the ramp magnitude for different spatial aggregations. For all the four ramp definitions, the distribution peak of the single site case has a relatively smaller probability density than multiple sites cases. This observation is more evident with the ramp Definition 3 in Fig. 9(c). For all spatial aggregations, at the peak of the distribution, both WFIP and STWPF present larger ramp magnitude comparing to the actual wind power ramps.



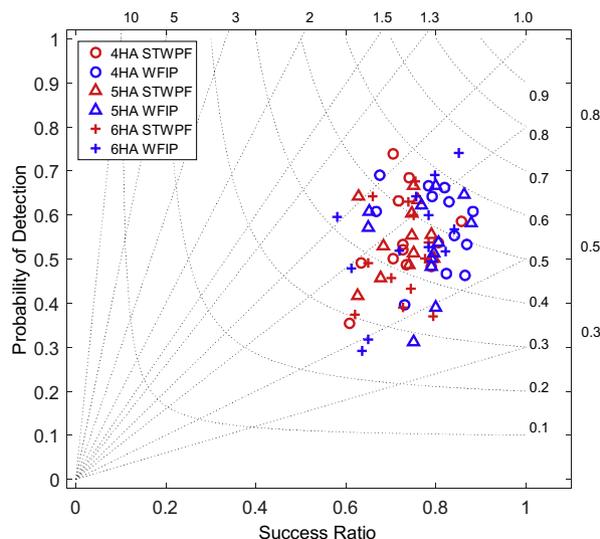
(a) Ramp Definition 1: Ramp magnitude only



(b) Ramp Definition 2: Ramp magnitude and duration



(c) Ramp Definition 3: Ramp rate



(d) Ramp Definition 4: Ramp direction, magnitude, and duration

Fig. 12. Comparison of monthly ramp forecasting performance of STWPF to WFIP for different forecasting timescales and ramp definitions.

4.3. Seasonal ramp forecasting performance

In addition to the annual ramp forecasting comparison, the seasonal ramp forecasting performance is also compared among different forecasting timescales for all wind power plants, as illustrated in Fig. 10. It is seen from all the four significant ramp definitions in Fig. 10(a)–(d) that, the ramp forecasting performs relatively better in fall and worse in summer for both the WFIP and STWPF. This can be partially attributed to the features responsible for ramps. The summer tends to be more convective (mesoscale); hence, it is more challenging to forecast (especially at the 4HA timescale) than the larger synoptic scale (such as fronts) features that cause ramps in the colder seasons. It is observed in Fig. 10(a) that during summer the ramp forecasting based on WFIP performs significantly better than that based on STWPF. However, during fall the ramp forecasting based on STWPF performs slightly better. This significant improvement of ramp forecasting performance in

summer based on the improved wind forecasts could play an important part in enhancing power system economics and reliability, considering that high electric demand is generally expected in ERCOT during the summer period. After comparing the ramp forecasting performances among the four different significant ramp definitions, it is evident that the WFIP ramp forecasts generally have a larger success ratio, POD, and CSI values with Definition 4, as shown in Fig. 10(d). The FBIAIS value is also closer to one with Definition 4.

Fig. 11 shows the distribution of the ramp magnitude for different seasons. It is seen that for all four seasons, at the peak of the distribution, both WFIP and STWPF present larger ramp magnitude comparing to the actual wind power ramps. It is also observed that distribution curves of the summer case are more fluctuated compared to other seasons. This again illustrates that ramp forecasting is relatively more challenging in summer.

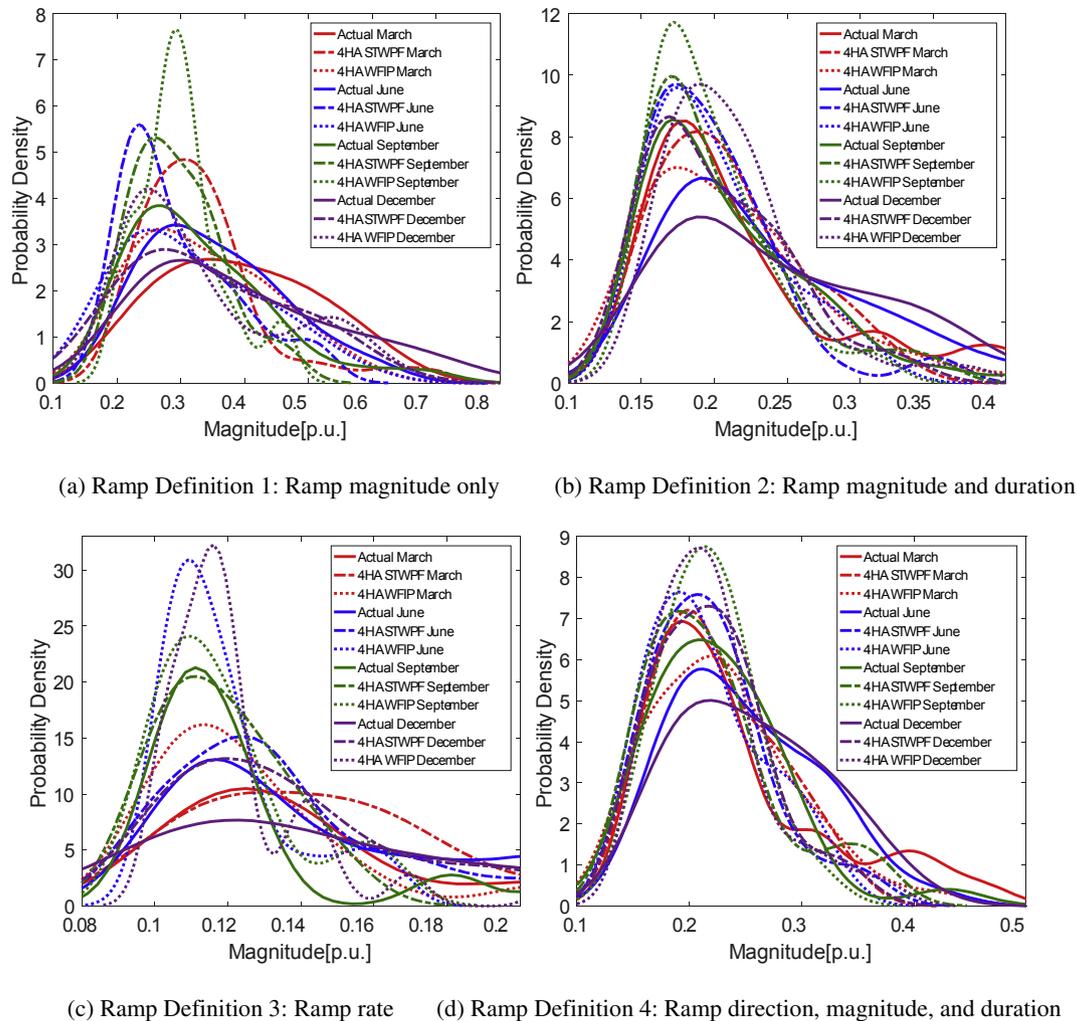


Fig. 13. Comparing monthly distribution of ramp magnitude between STWPF and WFIP.

4.4. Monthly ramp forecasting performance

Fig. 12 compares monthly ramp forecasting performance among the 4HA to 6HA STWPF and WFIP forecasts for all wind power plants within the WFIP southern study region. The 12 points represented by each symbol (e.g., circle) indicate the ramp forecasting performance in each month. It is observed that with different forecasting timescales and months, the variation in the ramp forecasting performance based on the significant ramp Definitions 2 and 4 is relatively less than that based on the ramp Definitions 1 and 3. Although the annual ramp forecasting tends to underforecast (FBIAS < 1 as shown in Fig. 6), the system tends to overforecast in a few months as shown in Fig. 12 in the cases of FBIAS > 1.

Fig. 13 shows the distribution of the ramp magnitude at four typical months (i.e., March, June, September, and December). It is seen that distribution curves in the September and December cases are more fluctuated, especially at the right tail of the distribution with Ramp Definition 3. We also found that most of the WFIP and STWPF distribution peaks are on the left side of the actual distribution peak, which again validate the underforecasting tendency.

4.5. Results discussion

The annual, seasonal, and monthly ramp forecasting evaluation results were analyzed. Overall, the ramp forecasting was improved

as a result of the improved wind power forecasts. It was found from the seasonal analysis that, the ramp forecasting performed relatively better in fall and worse in summer for both the WFIP and STWPF. This could be partially attributed to the features responsible for ramps. The summer tends to be more convective (mesoscale); hence, it is more challenging to forecast (especially at the 4HA timescale) than the larger synoptic scale (such as fronts) features that cause ramps in the colder seasons. We found from the multiple spatial scales analysis that, the number of significant ramps decreased when aggregating more wind power plants, which showed the smoothing effect from geographic diversity in wind power ramps. By comparing the four significant ramp definitions, we found that the performance of ramp forecasting was highly sensitive to the significant ramp definition. Therefore, it is important for power system operators to make multiple strategies to handle different types of wind power ramps based on their system characteristics, such as the flexibility capability, wind penetration level, etc.

5. Conclusion

This paper characterized the ramp forecasting performance of two forecasting systems: the experimental forecasts from the Wind Forecast Improvement Project (WFIP) and the current Electric Reliability Council of Texas (ERCOT) short-term wind power

forecast (STWPF). The WFIP experimental forecast system consists of an ensemble of high-resolution rapid-update numerical weather prediction (NWP) models in conjunction with a model output statistics process. A suite of statistical metrics was used to evaluate the overall improvement of the WFIP in short-term wind power forecasting accuracy at different forecasting horizons. Statistical analyses of the results showed that in most seasons and months the experimental WFIP provided a better performance than the current STWPF at all forecasting horizons.

An evaluation of the ramp forecasting improvement was performed based on four types of significant ramp definitions, which were determined based on the power change magnitude, direction, and duration. The wind power ramps were extracted using the optimized swinging door algorithm (OpSDA). Results showed that the OpSDA successfully captured wind power ramps for different definitions in terms of capacity and temporal differences. The ramp extraction results also found that the ramp magnitude varied more when the duration of the ramps was relatively short.

The ramp forecasting comparison results showed that enhanced WFIP forecasts improved the accuracy of wind power ramp forecasting, especially during the summer period in ERCOT. The results also showed that the ramp forecasting for both the WFIP and STWPF tended to underforecast during the whole year. The ramp forecasting performed relatively better with the ramp definition based on all three attributes: the power change direction, magnitude, and duration of the ramps.

Deterministic ramp forecasts are not able to predict the likelihood of occurrence or likelihood of different ramp event scenarios. An alternative to deterministic ramp forecasts is probabilistic ramp forecasts, which could provide a probability distribution of ramps or ramp rates. Future work will (i) develop probabilistic wind power ramp forecasts through NWP ensembles and (ii) evaluate the impacts of improved short-term wind power forecasting on probabilistic ramp forecasts.

Acknowledgments

This work was supported by the National Renewable Energy Laboratory under Subcontract No. XGJ-6-62183-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308).

References

- [1] EIA. *Electric power monthly*. U.S. Department of Energy; 2015.
- [2] ERCOT. Electric Reliability Council of Texas. Available: <http://www.ercot.com/>.
- [3] XcelEnergy. (2015). Available: <https://www.xcelenergy.com/>.
- [4] Lew D, Brinkman G, Ibanez E, Hodge B, King J. The western wind and solar integration study phase 2. Contract 2013;303:275–3000.
- [5] Wang J, Botterud A, Bessa R, Keko H, Carvalho L, Issicaba D, et al. "Wind power forecasting uncertainty and unit commitment. *Appl Energy* 2011;88:4014–23.
- [6] Alessandrini S, Sperati S, Pinson P. A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. *Appl Energy* 2013;107:271–80.
- [7] ERCOT. Report on the capacity, demand, and reserves in the ERCOT region, 2015–2024. Electric Reliability Council of Texas; 2014.
- [8] Ela E, Kirby B. ERCOT event on February 26, 2008: lessons learned. National Renewable Energy Laboratory; 2008.
- [9] CAISO. What the duck curve tells us about managing a green grid. California Independent System Operator; 2013.
- [10] Mai T, Wiser R, Sandor D, Brinkman G, Heath G, Denholm P, et al. Exploration of high-penetration renewable electricity futures. Vol. 1 of renewable electricity futures study. Golden, CO: National Renewable Energy Laboratory; 2012. Tech. Rep. NREL/TP-6A20-52409-1.
- [11] Gallego-Castillo C, Cuerva-Tejero A, Lopez-Garcia O. A review on the recent history of wind power ramp forecasting. *Renew Sustain Energy Rev* 2015;52:1148–57.
- [12] Ferreira C, Gama J, Matias L, Botterud A, Wang J. A survey on wind power ramp forecasting. Argonne National Laboratory (ANL); 2011.
- [13] Greaves B, Collins J, Parkes J, Tindal A. Temporal forecast uncertainty for ramp events. *Wind Eng* 2009;33:309–19.
- [14] Zheng H, Kusiak A. Prediction of wind farm power ramp rates: a data-mining approach. *J Sol Energy Eng* 2009;131:031011.
- [15] Potter CW, Gritmit E, Nijssen B. Potential benefits of a dedicated probabilistic ramp event forecast tool. In: Power systems conference and exposition, 2009. PSCE'09. IEEE/PES; 2009. p. 1–5.
- [16] Truewind A. AWS Truewind's final report for the Alberta forecasting pilot project. Wind Power Forecast PILOT Proj 2008. available: http://www.uwig.org/Alberta_PP_Final_Report_AWST_Jun25.pdf.
- [17] Freedman J, Markus M, Penc R. Analysis of West Texas wind plant ramp-up and ramp-down events. AWS Truewind, LLC; 2008.
- [18] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. "Current methods and advances in forecasting of wind power generation. *Renew Energy* 2012;37:1–8.
- [19] Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G. Wind power forecasting: state-of-the-art 2009. Argonne National Laboratory (ANL); 2009.
- [20] Botterud A, Wang J, Miranda V, Bessa RJ. Wind power forecasting in US electricity markets. *Electr J* 2010;23:71–82.
- [21] Wilczak J, Marquis M, Finley C, Freedman J. The Wind Forecast Improvement Project (WFIP): a public/private partnership for improving short term wind energy forecasts and quantifying the benefits of utility operations. Washington, DC: US Department of Energy (DOE); 2014. Tech. Rep. DE-EE0003080.
- [22] Freedman JM, Manobianco J, Schroeder J, Ancell B, Brewster K, Basu S, 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, final report. Albany, NY (United States): AWS Truepower, LLC; 2014.
- [23] Xue M, Droegemeier KK, Wong V. The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: model dynamics and verification. *Meteorol. Atmos Phys* 2000;75:161–93.
- [24] Xue M, Droegemeier KK, Wong V, Shapiro A, Brewster K, Carr F, et al. The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: model physics and applications. *Meteorol. Atmos Phys* 2001;76:143–65.
- [25] Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, et al. A description of the advanced research WRF version 2. 2005. DTIC Document.
- [26] Manobianco J, Zack JW, Taylor GE. In: Workstation-based real-time mesoscale modeling designed for weather support to operations at the Kennedy space center and cape canaveral Air station; 1996.
- [27] Zhang J, Florita A, Hodge B-M, Freedman J. Ramp forecasting performance from improved short-term wind power forecasting. In: ASME 2014 international design engineering technical conferences and computers and information in engineering conference; 2014.
- [28] Cui M, Zhang J, Florita AR, Hodge B-M, Ke D, Sun Y. An optimized swinging door algorithm for wind power ramp event detection. In: IEEE power & energy society general meeting, Denver, CO; 2015.
- [29] Cui M, Zhang J, Florita AR, Hodge B-M, Ke D, Sun Y. An optimized swinging door algorithm for identifying wind ramping events. *IEEE Trans Sustain Energy* 2016;7:150–62.
- [30] Roebber PJ. Visualizing multiple measures of forecast quality. *Weather Forecast* 2009;24:601–8.