A WRF Ensemble for Improved Wind Speed Forecasts at Turbine Height

ADAM J. DEPPE AND WILLIAM A. GALLUS JR.

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

EUGENE S. TAKLE

Department of Geological and Atmospheric Sciences, and Department of Agronomy, Iowa State University, Ames, Iowa

(Manuscript received 2 October 2011, in final form 15 May 2013)

ABSTRACT

The Weather Research and Forecasting Model (WRF) with 10-km horizontal grid spacing was used to explore improvements in wind speed forecasts at a typical wind turbine hub height (80 m). An ensemble consisting of WRF model simulations with different planetary boundary layer (PBL) schemes showed little spread among the individual ensemble members for forecasting wind speed. A second configuration using three random perturbations of the Global Forecast System model produced more spread in the wind speed forecasts, but the ensemble mean possessed a higher mean absolute error (MAE). A third ensemble of different initialization times showed larger model spread, but model MAE was not compromised. In addition, postprocessing techniques such as training of the model for the day 2 forecast based on day 1 results and bias correction based on observed wind direction are examined. Ramp event forecasting was also explored. An event was considered to be a ramp event if the change in wind power was 50% or more of total capacity in either 4 or 2 h or less. This was approximated using a typical wind turbine power curve such that any wind speed increase or decrease of more than 3 m s^{-1} within the 6–12 m s^{-1} window (where power production varies greatly) in 4 h or less would be considered a ramp. Model MAE, climatology of ramp events, and causes were examined. All PBL schemes examined predicted fewer ramp events compared to the observations, and model forecasts for ramps in general were poor.

1. Introduction

In recent years, wind energy production has undergone rapid growth, and the U.S. Department of Energy goal of having 20% of the nation's electrical energy from wind by 2030 will require continued growth (Department of Energy 2008). Wind, unlike other sources of energy, varies substantially over both space and time. Therefore, the production rates of wind energy fluctuate more strongly than those of other traditional fossil fuel sources of energy generation. To optimize wind for power generation, accurate forecasts are needed.

Unfortunately, there have been few evaluations of model forecasts of winds at 80 m, a height where the influence of turbulent fluxes of momentum, heat, and moisture from the earth's surface can vary greatly depending on the time of day, season, and vertical temperature stratification of the atmosphere. Meteorologists traditionally have focused wind forecasts at the 10-m level, a height at which official wind observations are routinely taken and a level at which winds are strongly influenced by surface friction. Prior wind forecasting research in the western United States has focused on flow in complex terrain (e.g., Wood 2000; Ayotte et al. 2001) and is therefore not applicable in Iowa where boundary layer stratification, low-level jets (LLJs), and changing surface conditions are likely to be the dominant factors providing uncertainty in short-term forecasts at 80 m. Other modeling studies have taken a more statistical approach to predicting wind speed at different levels (Huang and Chalabi 1996); however, none have been reported for the state of Iowa, despite it being the state with the largest percentage of total power per capita coming from wind energy in 2010 (Department of Energy 2010). Even fewer studies have examined the forecasting of ramp events, defined as rapid changes in wind speed that lead to extreme changes in wind power output. Large ramp events causing a 50% or greater change in the capacity of the

Corresponding author address: Adam J. Deppe, 741 Hawkmount Circle, Apt. B, Chesterfield, MO 63017.
E-mail: ajdeppe@gmail.com

DOI: 10.1175/WAF-D-11-00112.1

wind speed were found to occur less than 7% of the time within 4 h in a United Kingdom wind farm study (Greaves et al. 2009) and less than 4% of the time within 2 h in a California wind farm study (Zack 2007). Although rare, ramp events occurring between the cut-in speed (wind speed when turbine begins to generate power) and the rated wind speed (wind speed when the turbine begins to generate its designated rated power; see Fig. 3 for an example) are extremely costly to energy companies because they may cause blackouts and overload the grid (Francis 2008). Along with being rare, ramp events are also difficult to forecast. It was found that ramp events were captured less than 36% of the time by a private forecast company forecasting for six wind farms in the United States (Greaves et al. 2009).

In the present study, the ability of version 3.1.1 of the Weather Research and Forecasting Model (WRF) to accurately reproduce 80-m wind speeds and ramp events was evaluated by comparing WRF simulations using six different planetary boundary layer (PBL) schemes to observations of 80-m wind speed gathered at the Pomeroy, Iowa wind farm site. The sensitivities of the two most widely used PBL schemes—the Yonsei University scheme (YSU) and the Mellor–Yamada–Janjić scheme (MYJ)—along with the quasi-normal scale elimination PBL scheme (QNSE), the Mellor–Yamada Nakanishi and Niino level-2.5 PBL scheme (MYNN 2.5), the MYNN level 3.0 PBL scheme (MYNN 3.0), and the Pleim PBL scheme [also called the asymmetric convective model [ACM2)] are examined. A brief review of the six different schemes can be found in the appendix.

2. Model configuration and data

For most of the simulations examined in the present study, a single domain with 10-km horizontal grid spacing was used, although some tests were performed embedding a nested 4-km grid-spacing domain within the coarser domain (Fig. 1). Both domains had 47 vertical levels, with 16 levels in the lowest 1300 m and an average vertical spacing of around 15 m in the lowest 100 m. The lowest half-sigma levels (heights) on which velocity data were present were 1.0 (surface), 0.9995 (10 m), 0.998 (25 m), 0.996 (40 m), 0.994 (56 m), 0.992 (72 m), 0.990 (88 m), 0.988 (108 m), 0.985 (137 m), 0.9805 (180 m), 0.974 (244 m), 0.962 (377 m), 0.944 (546 m), 0.9215 (761 m), 0.8945 (1016 m), and 0.8645 (1300 m). The physical schemes used include Ferrier microphysics (Ferrier et al. 2002), the Rapid Radiation Transfer Model (RRTM; Mlawer et al. 1997) for longwave radiation, and the Dudhia scheme (Dudhia 1989) for shortwave radiation. The Noah land surface scheme (Ek et al. 2003) was used for all of the model runs except for the one using the revised Asymmetric Convective Model (ACM2) scheme, which employed the Pleim–Xiu scheme (Pleim and Xiu 1995; Xiu and Pleim 2001) since the Noah scheme was not applicable with the ACM2 PBL scheme in the version of WRF used. A cumulus scheme was not used for the 4-km runs, while the 10-km runs used the Kain–Fritsch model (Kain 2004). Six different PBL and surface layer schemes were evaluated in this study. The MYJ PBL scheme simulation used the Janjić Eta Model Monin–Obukhov surface layer scheme, the MYNN 2.5 and MYNN 3.0 PBL schemes used the MYNN surface layer scheme, the ACM2 PBL scheme used the Pleim–Xiu surface layer scheme, the QNSE PBL scheme used the QNSE surface layer scheme, and the YSU PBL scheme used the Monin–Obukhov (Hong and Pan 1996) surface layer scheme.

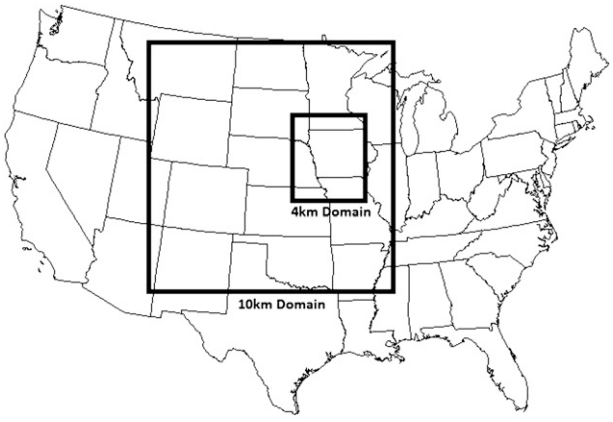

FIG. 1. The 10- and 4-km model domains used in this study.

The 54-h model runs were initiated at 1800 UTC [1200 local standard time (LST)], 0000 UTC (1800 LST), and 0600 UTC (0000 LST) using both 1° horizontal grid spacing Global Forecast System (GFS) output, and 12-km horizontal grid-spacing North American Model (NAM) output for initial and lateral boundary conditions (ILBCs). For each 2-day period, 12 forecasts were made, one for each PBL scheme and each ILBC (Table 1). Because energy companies often use different techniques for nowcasting of winds in the very short term compared to forecasting during the 1–2-day period, we have chosen to consider the first 6 h of each model run as a spinup period, and have not included this period in the evaluations of error. It should be noted that mean absolute errors (MAEs) in model runs using four of the six PBL schemes tested were somewhat higher during the first 6 h than over the following 48, but differences were usually less than 5%. In addition, a case study was performed where instantaneous wind output was compared to hourly averaged model wind output, and no significant differences were noted; therefore, hourly

TABLE 1. Parameterization combinations used to create ensemble members.

Member No.	PBL scheme	Land surface scheme	Land layer scheme	Initial boundary conditions
1	YSU	Noah	Monin–Obukhov	GFS
2	MYJ	Noah	Janjić Eta Model Monin–Obukhov	GFS
3	QNSE	Noah	QNSE	GFS
4	MYNN 2.5	Noah	MYNN	GFS
5	MYNN 3.0	Noah	MYNN	GFS
6	ACM2	Pleim–Xiu	Pleim–Xiu	GFS
7	YSU	Noah	Monin–Obukhov	NAM
8	MYJ	Noah	Janjić Eta Model Monin–Obukhov	NAM
9	QNSE	Noah	QNSE	NAM
10	MYNN 2.5	Noah	MYNN	NAM
11	MYNN 3.0	Noah	MYNN	NAM
12	ACM2	Pleim–Xiu	Pleim–Xiu	NAM

averaged model output was used in this study. Observed data for comparison with model results were taken from an 80-m meteorological tower on the southwest side of the Pomeroy wind farm (Fig. 2) at 10-min increments and averaged over 1-h periods centered on each hour to measure the true sustained wind speed. The 80-m wind speeds were evaluated from June 2008 through September 2010, excluding periods when data were missing, while 58 cases spanning 116 days from June 2008 through June 2009 were evaluated in the wind ramp portion of this paper.

3. Methodology

Two forecast evaluations were performed. The first used MAE and bias to evaluate wind speed forecasts at 80-m elevation. For this examination, an operational ensemble was developed based on the MAE of numerous member configurations examined in three sets of tests. The first set of tests, prerun modification, explored different time initializations, grid spacing, and perturbations of the GFS ILBCs. The second set of tests, postprocessing, focused on three techniques, the neighborhood approach, training of the model, and bias correction. In the neighborhood approach, forecast values at grid points around the validation tower were averaged in lieu of using the grid point closest to the tower. The neighborhood approach has been successfully used to improve precipitation forecasting (Theis et al. 2005; Ebert 2009), although results have not been reported when applied to wind speed forecasting. The second technique examined training of the model based on model MAEs in the first 24-h period. The three members with the lowest MAE during the first 24-h period were used to form an ensemble to forecast the 24–48-h period, referred to hereafter as day 2. The third technique focused on bias corrections based on (i) wind speed, (ii) wind direction, (iii) wind speed and direction, and (iv) the diurnal cycle. From the results of both prerun and postprocessing tests, a final ensemble to be used operationally was developed, hereafter known as final OP. The Wilcoxon signed-rank test was used to determine if the improvements in the final OP ensemble were significant. The Wilcoxon signed-rank test was chosen as this test does not depend

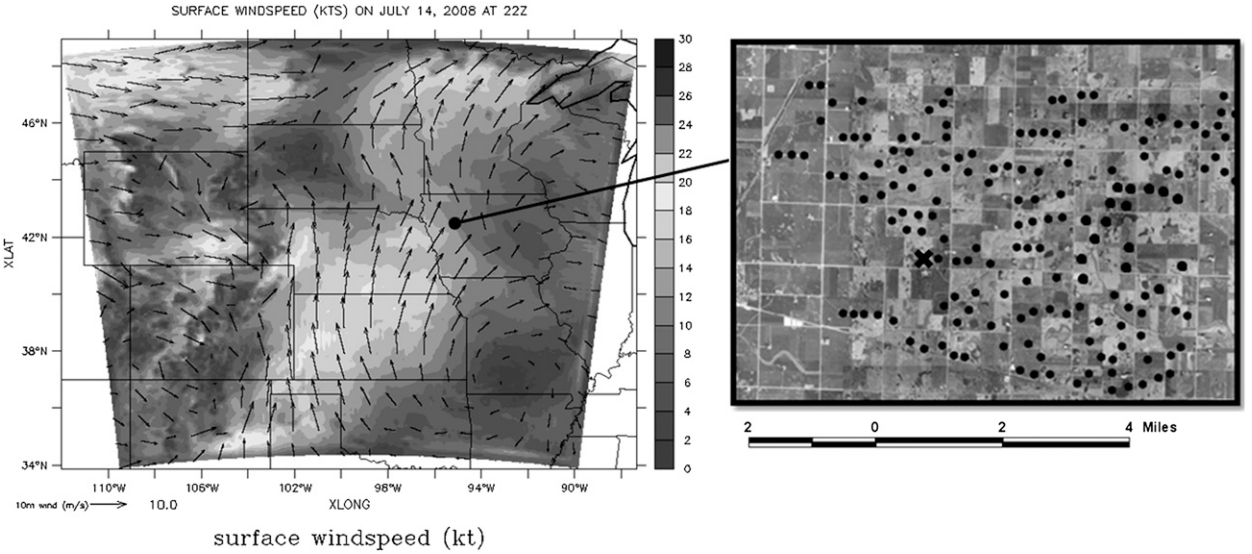

FIG. 2. (left) The 10-km domain with (right) outline of the Pomeroy wind farm where the individual wind turbines are the black dots and the 80-m meteorological tower (observed data location) is denoted with the ×.

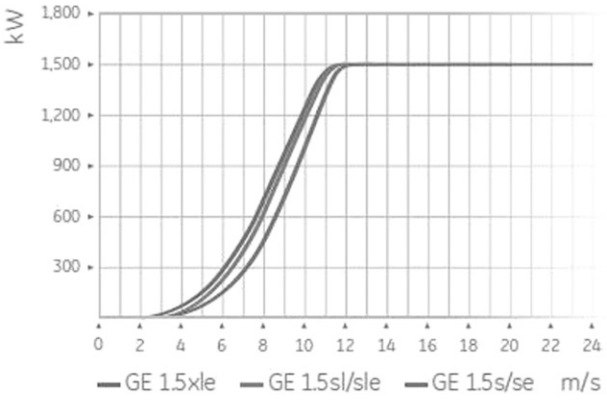

FIG. 3. Power curve for the 1.5-MW wind turbines used at the Pomeroy wind farm. Cut-in speed is around 3.5 m s^{-1} while the rated wind speed is around 12 m s^{-1} (General Electric Company 2005).

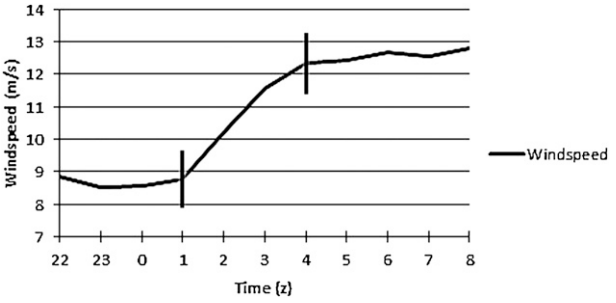

FIG. 4. Example of a ramp-up event. Start is assumed to be at 0100 UTC when sharp change in wind speed begins and ends when the change in wind speed becomes minimal at 0400 UTC.

on the distribution of the data and is resistant to outliers (Wilks 2006, 160–162). Although we focus on MAE in the results that follow, it is important to note that the use of root-mean-square error (RMSE) might lead to different results as larger errors would be penalized more.

The second test of forecast accuracy focused on ramp events at 80 m. In this paper, an event was considered to be a ramp if the change in wind power over 4 h or less was 50% or more of total capacity (Greaves et al. 2009). We also looked at results if this window was halved to 2 h. The change in power was approximated using a typical wind turbine power curve (Fig. 3). Contained by 6–12 m s^{-1} (an area where power production varies greatly), any wind speed increase or decrease of more than 3 m s^{-1} in 4 h or less was considered a ramp (similar to Greaves et al. 2009). Defining the start and end of a ramp event was somewhat subjective. The start of a ramp event was defined as a sharp change in the wind speed while the end was marked by a minimal change in the wind speed (Fig. 4). Ramps were classified into two categories: ramp-ups (increase in speed within 4 or 2 h) and ramp-downs (decrease in speed within 4 or 2 h), similar to the technique used by Freedman et al. (2008) for surface data in a west Texas study. Wind observations were put through extensive quality control, and cases were chosen from the subset of days when reliable data existed. The wind data archive contained wind speed values every 10 min, and observed ramps were determined using both the 10-min data and top-of-the-hour data. The results to follow focus on the hourly data, since the large set of model output only had a temporal frequency of 1 h. Model accuracy was evaluated in three areas: number of ramp events forecasted, frequency of events, and model error. MAE, probability of detection (POD), false alarm rate (FAR), and threat score (TS) were calculated to determine model accuracy using the following equations:

$$\mathrm{POD} = \frac{\text{Total number of correct event forecasts (Hits)}}{\text{Total number of events observed}}, \tag{1}$$

$$\mathrm{FAR} = \frac{\text{Total number of false alarms}}{\text{Total number of events forecasted}}, \quad \text{and} \tag{2}$$

$$\mathrm{TS} = \frac{\text{Total number of correct event forecasts (Hits)}}{\text{Total number of events forecasted} + \text{Number of misses}}. \tag{3}$$

4. Evaluation of 80-m wind forecasts

To compare the 6 different PBL schemes, 32 cases (8 from each season; winter, spring, summer, and fall) were chosen at random (using a random number generator) during periods having high quality observed data. We created an ensemble from WRF model runs with different PBL schemes at the same initialization time, based on results of Harrison et al. (1999) and Stensrud et al. (2000), who found that varying the model physics was a powerful method for creating a forecast ensemble. However, in our study changing the PBL schemes produced little ensemble member spread among all six PBL schemes using the same initialization time

TABLE 2. MAE (m s^{-1}) associated with six PBL schemes using the 0000 UTC time initialization and the GFS ILBCs from 10 cases during January 2010. The six-member ensemble average and the standard deviation (measure of model spread) are also listed.

PBL scheme	MAE	Std dev
MYJ	1.38	—
MYNN 2.5	1.43	—
MYNN 3.0	1.38	—
Pleim	1.29	—
QNSE	1.39	—
YSU	1.31	—
Ensemble	1.26	0.66

TABLE 3. MAE (m s^{-1}) associated with 3 different GFS perturbations using the YSU and MYNN3.0 PBL schemes from 10 cases during January 2010. The two-member ensemble average and the standard deviation (measure of model spread) are also listed.

Perturbation No.	2	4	15	Std dev
MYNN 3.0 MAE	1.88	1.73	1.80	—
YSU MAE	1.60	1.59	1.72	—
Ensemble MAE	1.58	1.53	1.62	0.98

(Table 2). Small spread is good if all model versions are predicting speeds correctly; however, more often it results in all models yielding incorrect forecasts (Houtekamer 1993; Whitaker and Loughe 1998). Therefore, three techniques were investigated to improve model scheme spread and MAE over 10 cases during January 2010, for which GFS perturbation data were available. It should be noted that extensive snow cover was present at the site during January 2010, and the substantial impacts snow cover has on near-surface processes could lead to different model performance during this period than at other times of the year. For the prerun modifications, only the YSU and MYNN 3.0 PBL schemes were used to reduce computational expenses. These schemes were selected because they often led to relatively different results.

a. Prerun modification

The first attempt to improve model MAE used different perturbations of the initial and lateral boundary conditions for the GFS model. As of 20 May 2006, GFS perturbations were developed using an ensemble transform (ET) technique (Wei et al. 2006). ET replaced the breeding method and eliminated paired perturbations, making all perturbations random to each other. Therefore, in this study we selected three perturbation members (2, 4, and 15) to compare against the three initialization times tested later. The perturbations picked were run for 10 cases in January 2010 using the YSU and MYNN 3.0 PBL schemes. The results of this trial increased the model spread; however, model MAE also increased (Table 3) from the six PBL schemes tested at 0000 UTC (Table 2).

The second approach for improving model MAE changed the grid spacing. A two-member ensemble using the 10-km grid and the YSU and MYNN 3.0 PBL schemes was created and evaluated against a two-member ensemble using a 4-km grid and the YSU and MYNN 3.0 PBL schemes over a 10-day period during January 2010. Both the YSU and MYNN 3.0 simulations, and the ensemble mean, showed lower MAEs with 10-km grid spacing compared to 4 km (Table 4), although the Wilcoxon signed-rank test showed the results were not highly significant. Another goal of this study was to design an ensemble that could be used by wind energy companies. With computing power limited in most private companies, running 10-km model simulations is much more feasible than running 4-km cases. Therefore, because the MAE of the 4-km runs was not better than that of the 10-km runs, we focused the remainder of our study on simulations using 10-km grid spacing.

The third approach to improving the model MAE changed the time of initialization. The motivation for testing different time initializations or time-lagged ensembles came from the success and usefulness achieved in many other previous short- to medium-range forecasting studies (Hoffman and Kalnay 1983; Dalcher et al. 1988; Walser et al. 2004; Lu et al. 2007). In our study, WRF simulations using the YSU and MYNN 3.0 PBL schemes were initialized at 1800 UTC (1200 LST), 0000 UTC (1800 LST), and 0600 UTC (0000 LST) over a 10-day period in January 2010. The 0000 UTC (1800 LST) and 1800 UTC (1200 LST) time initializations showed the lowest MAEs while the 0600 UTC (0000 LST) initialization, the initialization closest to the forecast period, showed the highest MAE (Table 5), although these results were not highly significant. To better determine when 0600 UTC (0000 LST) initialization run errors were largest, a 42-h time series of wind speed, averaged over 10 different cases in January 2010, was compared to observational data in Fig. 5a. The YSU scheme largely underpredicted the wind speed from forecast hour 18 to 32 (mostly nighttime) (Fig. 5a), resulting in a high MAE (Fig. 5b) during this period. However, compared to the perturbation ensemble, the time initialization

TABLE 4. MAE (m s^{-1}) associated with the wind speed at 80 m from 2 different grid spacings (4 and 10 km) from 10 cases during January 2010. The two-member ensemble average is also listed.

Grid spacing (km)	MYNN 3.0 MAE	YSU MAE	Ensemble MAE
10	1.37	1.29	1.18
4	1.70	1.33	1.27

TABLE 5. MAE (m s^{-1}) associated with the wind speed at 80 m from three different initialization times from 10 cases during January 2010. The two-member ensemble average and the standard deviation (measure of model spread) are also listed.

Time initialization (UTC)	1800	0000	0600	Std dev
MYNN 3.0 MAE	1.42	1.37	1.38	—
YSU MAE	1.32	1.29	1.61	—
Ensemble MAE	1.23	1.18	1.28	1.09

ensemble showed higher model spread with a highly significant lower MAE. Therefore, our final OP ensemble was designed to include both different PBL schemes and 0000 UTC (1800 LST) and 1800 UTC (1200 LST) time initializations.

b. Postprocessing

We investigated three postprocessing techniques, which included training of the model, the neighborhood approach, and bias correction of the wind speed. The first postprocessing technique trained the model based on day 1 results. In this method, day 1 forecasts (hours 6–30) were analyzed and the three most accurate PBL schemes (lowest MAEs) were chosen and a selected ensemble was developed to forecast day 2 wind speeds. The three least accurate PBL schemes (highest MAEs) were chosen as members of the nonselected ensemble (Table 6). The most accurate day 1 forecasts were not found to always result in the most accurate day 2 forecasts. From the 15 cases studied, the nonselected ensemble showed the lowest MAE 4 out of the 15 times (27%), the selected ensemble showed the lowest MAE 5 out of the 15 times (33%), and the ensemble incorporating all six model members showed the lowest MAE 6 out of the 15 times (40%). Therefore, the training approach was not a reliable method for predicting wind speed as conditions change too much from day to day, a result similar to that found in Briggs and Ruppert (2004) and Hall et al. (2010).

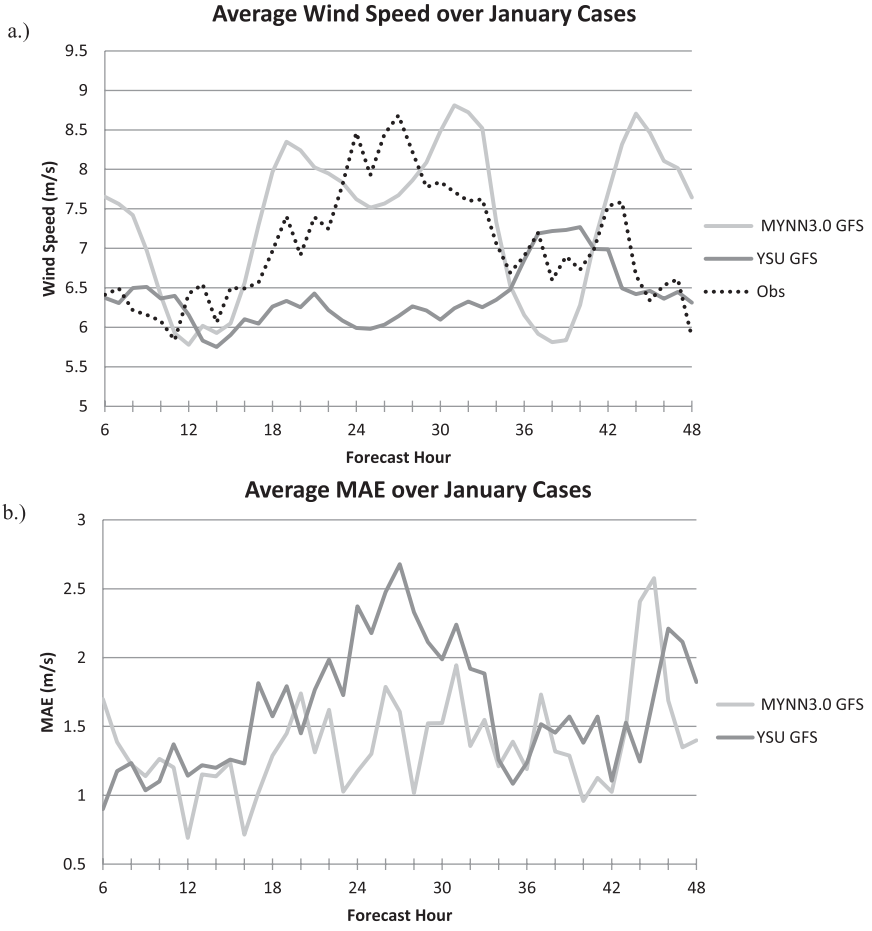

FIG. 5. A 48-h time series (first 6 h not shown due to model "start up") of (a) wind speed and (b) MAE, averaged over 10 different cases during January 2010. Model runs using 0600 UTC (0000 LST) time initializations and MYNN3.0 (light gray) and YSU (dark gray) PBL schemes with GFS initial and lateral boundary data. Observations in (a) are shown with the dotted line.

TABLE 6. MAE (m s^{-1}) calculated for the first 24-h period. The 3 PBL schemes with the lowest MAEs were chosen, making up the day 2 selected ensemble. Times selected indicates the number of times a model was chosen as a member of the day 2 selected ensemble. The nonselected ensemble incorporated the least accurate models for the first 24-h period. Day 2 all-member ensemble incorporated all six model members.

Model Number	Day 1 MAE (m s^{-1})	Times selected
00 UTC MYJ GFS with a 10-km grid spacing	2.51	5
00 UTC MYJ NAM with a 10-km grid spacing	2.61	6
00 UTC Pleim NAM with a 10-km grid spacing	2.58	4
00 UTC Pleim GFS with a 10-km grid spacing	2.36	9
00 UTC YSU NAM with a 10-km grid spacing	2.32	11
00 UTC YSU GFS with a 10-km grid spacing	2.37	10
Ensemble Mean	1.97	
Day 2 selected ensemble best MAE	Day 2 non-selected ensemble best MAE	Day 2 All-Member Ensemble best MAE
5/15	4/15	6/15

The second postprocessing technique used the neighborhood approach. Instead of basing a forecast for a location on the winds predicted at the model grid point closest to that location, in the neighborhood approach, a set of grid points around the location of interest was averaged. The results of this test varied for different PBL schemes. The YSU scheme, a nonlocal and first-order closure scheme, became more accurate when a large set of grid points around the location of interest was averaged while the MYNN 3.0 scheme, a local and second-order closure scheme, became less accurate when a large set of grid points around the location of interest was averaged (Table 7). The reason for this may be due to the larger variation in the surface layer length scale in the MYNN PBL scheme, which is a function of the stability parameter (z/L). If z/L varies significantly from grid point to grid point, the turbulent mixing will be more variable in the horizontal, making neighborhood grid points more independent. Ensemble results from the neighborhood approach show lower MAEs when an average was taken from a box consisting of 17×17 grid points; ensemble MAEs did not improve when averaging over larger areas (Table 7). However, the improvement was very small and not statistically significant. The improvement was also not as large as that resulting from the other methods tested.

The third postprocessing technique used biases observed in the PBL schemes to adjust the forecasts. A bias in the model was computed by analyzing 30 random cases (0000 UCT model initializations) from all seasons between June 2008 and June 2009 (Fig. 6) as a training set. Ideally, biases should be computed for each season since land surface characteristics, among other parameters, can differ greatly between seasons and potentially can lead to different model biases, but our sample size was too small to allow that. However, future studies should compare seasonal biases associated with wind speed.

In this training set of cases, all PBL schemes except the YSU exhibited a diurnal cycle in the bias. A negative bias, or underprediction of the wind speed, occurred between 1200 and 2000 UTC (0600 and 1400 LST), while a positive bias (overprediction) occurred from 2000 to 1200 UTC the next day (1400 to 0600 LST). The same pattern existed in day 2 of the 54-h forecast and was present in runs using both the GFS and NAM ILBCs. A similar result, although for surface wind speed, was found by Zhang and Zheng (2004). These consistent diurnal trends in model error allowed for bias correction of the forecasts.

Four bias-correction approaches were examined: one based on the diurnal cycle of wind speeds, one using wind speed along with direction, one based on wind speed only, and the final one using wind direction only. The diurnal cycle bias was computed as the mean bias of the wind speed as a function of the hour of the day. The wind-speed-only bias correction was computed as the mean bias of the wind speed as a function of the wind speed range (i.e., 3–6 m s^{-1}). The direction-only bias was computed in a similar way, only using a range of degrees instead of wind speed. And last, the wind-speed-and-direction bias was computed as the mean bias of the wind speed as a function of both the wind speed range and direction range. Please note that each of the bias

TABLE 7. MAE (m s^{-1}) for wind speed at 80 m associated with the neighborhood approach.

Grid averaging	MYNN3.0 MAE	YSU MAE	Ensemble MAE
Point	1.37	1.29	1.18
3×3	1.36	1.28	1.17
5×5	1.36	1.25	1.16
11×11	1.38	1.18	1.14
17×17	1.39	1.16	1.13
21×21	1.40	1.17	1.14

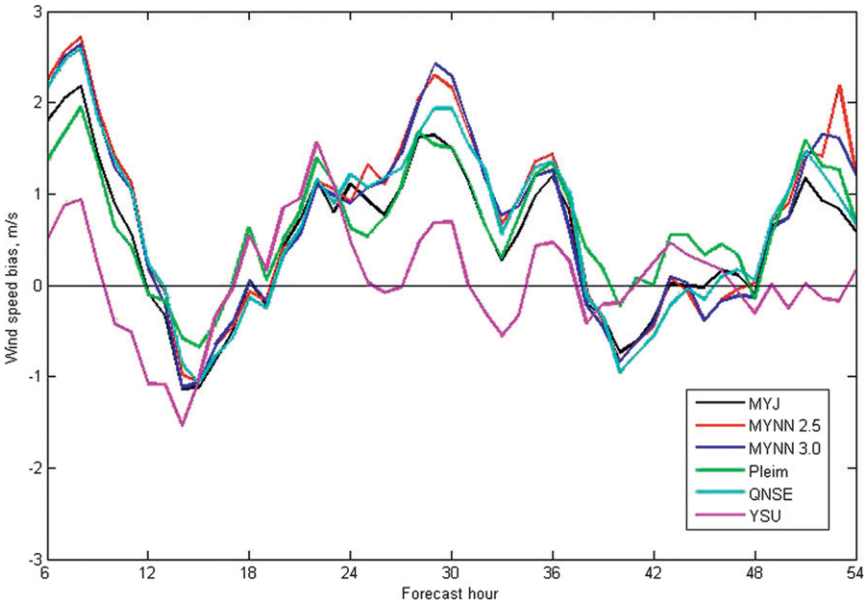

FIG. 6. Average PBL biases by hour using the 0000 UTC (1800 LST) initialization (first 6 h not shown due to model start up). Each line represents a different PBL scheme; MYJ (gray), MYNN 2.5 (red), MYNN 3.0 (blue), Pleim or ACM2 (green), QNSE (aqua), and YSU (magenta).

corrections above was applied hourly to the test set discussed below.

Based on the 4 bias correction approaches above, a test set using only the 0000 UTC (1800 LST) time initializations with GFS ILBCs over a 32-day period from 11 October to 11 November 2008 (Table 8) was used to determine which bias correction approach resulted in the lowest MAE. Please note that due to the expense associated with an ensemble of computational runs, only the 0000 UTC (1800 LST) time initialization, using GFS ILBCs, was evaluated in in this test. Forecasts using the wind speed bias correction showed the lowest MAE, and therefore this bias correction was selected to be used to improve the wind speed forecast in the final OP ensemble.

To determine the six members that would make up the final OP ensemble, a 15-day test period from 14 to 28 August 2009 was evaluated. New wind speed bias corrections appropriate for this mix of 0000 UTC (1800 LST) and 1800 UTC (1200 LST) time initializations and GFS and NAM ILBCs were determined using a training set of 30 cases over all four seasons, and these corrections were applied to this 15-day test period (Table 9). Please note that this test was done to confirm which combination of the 0000 UTC (1800 LST)–1800 UTC (1200 LST) time initializations and GFS–NAM ILBCs would result in the lowest MAEs. From a possible 24 different combinations, a six-member ensemble was created. The six members found to have the lowest MAEs after the completion of the two tests mentioned above included the

TABLE 8. MAEs (m s^{-1}) associated with different bias corrections developed for each PBL scheme for the 0000 UTC GFS ILBCs. This case study was done from 11 Oct 2008 to 11 Nov 2008.

Bias corrections	MYJ	MYNN 2.5	MYNN 3.0	Pleim	QNSE	YSU	Ensemble
No bias	2.34	2.49	2.41	2.36	2.45	2.28	2.27
Diurnal cycle	2.29	2.33	2.28	2.27	2.30	2.21	2.18
Wind direction	2.27	2.27	2.26	2.29	2.28	2.24	2.17
Wind speed and direction	2.15	2.16	2.14	2.17	2.17	2.10	2.05
Wind speed	2.05	2.04	2.01	2.09	2.07	1.99	1.97
Best	0.29	0.45	0.40	0.27	0.38	0.29	0.30
Improvement	—	Wind speed	Wind speed	Wind speed	—	—	—
	Wind speed				Wind speed	Wind speed	Wind Speed
Improvement (%)	14.1	22.1	20.0	13.0	18.4	14.6	15.2

TABLE 9. MAE (m s^{-1}) associated with different PBL schemes using the wind speed bias correction. The lowest MAE was produced by the YSU and Pleim schemes. The case study was done during 14–28 Aug 2009.

	MYJ	MYNN 2.5	MYNN 3.0	Pleim	QNSE	YSU	Ensemble
GFS 0000 UTC	1.59	1.66	1.66	1.52	1.65	1.57	1.48
GFS 1800 UTC	1.68	1.81	1.72	1.61	1.77	1.63	1.58
NAM 0000 UTC	1.67	1.71	1.69	1.63	1.71	1.57	1.56
NAM 1800 UTC	1.66	1.75	1.74	1.60	1.70	1.63	1.57

1800 UTC Pleim GFS, 1800 UTC Pleim NAM, 0000 UTC Pleim GFS, 0000 UTC YSU NAM, 0000 UTC YSU GFS, and 0000 UTC MYJ GFS (Table 10). All six members used 10-km grid spacing. From these six members, the final OP ensemble was created. Most of the time, simulations using GFS ILBCs showed lower MAEs than those using NAM, so four out of the six members of the final OP used the GFS ILBCs. Five out of the six members that formed the ensemble used either the Pleim or YSU PBL scheme. As noted previously, the YSU and the Pleim PBL schemes use first-order closure and nonlocal mixing, while the other four PBL schemes tested use TKE closures and involve local mixing. Therefore, from our results, it appears that nonlocal PBL schemes provide the lowest MAEs for 80-m wind speed forecasts in northwest Iowa.

To evaluate the final OP ensemble, a deterministic forecast—the 0000 UTC YSU GFS; the PBL scheme that showed the lowest MAE—as well as four other six-member ensembles, were compared. The standard deviation was also calculated to determine model spread and compared to the final OP ensemble. To test the final OP ensemble, 25 random cases from the summer and fall of 2010 were used. The six-member final OP ensemble model had the lowest MAEs of any of the other six-member ensembles tested, both before and after the wind speed bias correction (Table 11). Based on these results, significance testing was done using the Wilcoxon signed-rank test. When comparing the non-bias-corrected final OP ensemble to the other non-bias-corrected six-member ensembles and the deterministic forecast, the improvement in MAE of the final OP ensemble was significant, with p values all less than 0.08. This indicates that an ensemble consisting of different time initializations and the YSU, MYJ, and Pleim PBL schemes was more accurate than an ensemble constructed of all six PBL schemes. Finally, comparing the bias-corrected final OP ensemble model to the non-bias-corrected six-member ensembles and the deterministic forecast, the improvement in MAE of the bias-corrected final OP ensemble was highly significant, with all p values less than 0.004. This demonstrates that the final OP ensemble designed in this paper shows a significant degree of improvement in wind speed forecasting over the other approaches tested. The standard deviation of the final OP ensemble was also larger than that of any of the other ensembles, indicating a larger spread in the final OP ensemble, which should be helpful in capturing outlier events. In a perfectly calibrated ensemble, increased spread also would identify episodes of higher forecast uncertainty. However, the relationship between the spread and the RMSE is not linear in our ensemble (figure not shown), so that despite having larger spread than the other ensembles tested, the OP ensemble would not necessarily perform as well at forecasting the forecast skill as a perfectly calibrated ensemble would.

Although a 0.15 m s^{-1} improvement in MAE in some circumstances is small, in the area of wind energy, this may be a substantial improvement. In a 2008 Department of Energy (DOE) publication, it was stated that, ''given that a 1% error in wind speed estimates for a 100-MW wind generation facility can lead to losses approaching $12 000 000 over the lifetime of that plant, a better understanding of the physical and dynamic processes across the range of scales that create a particular wind climate is needed'' (Schreck et al. 2008, 81–82). The ensemble developed in this paper shows a 10% improvement in wind speed prediction, which, considering

TABLE 10. Parameterization combinations used in the final OP ensemble to forecast wind speed at 80 m.

Member No.	PBL scheme	Time initialization (UTC)	Land surface scheme	Land layer scheme	Initial boundary conditions
1	ACM2	1800	Pleim–Xiu	Pleim–Xiu	GFS
2	ACM2	1800	Pleim–Xiu	Pleim–Xiu	NAM
3	ACM2	0000	Pleim–Xiu	Pleim–Xiu	GFS
4	YSU	0000	Noah	Monin–Obukhov	NAM
5	YSU	0000	Noah	Monin–Obukhov	GFS
6	MYJ	0000	Noah	Janjić Eta Monin–Obukhov	GFS

TABLE 11. MAEs (m s^{-1}) of final OP ensemble after wind speed bias correction compared to other six-member ensembles tested for 25 cases during the summer and fall of 2010. The deterministic forecast is the best individual model found from the period studied. Standard deviation (measure of model spread) for each ensemble is also calculated. The boldface value indicates a high level of statistical improvement from the non-bias-corrected <0.1 determined from a Wilcoxon signed-rank test. The italicized value indicates a high level of statistical improvement from the non-bias-corrected six-member ensembles–deterministic forecast to the bias-corrected final OP ensemble, with p values <0.1 determined from a Wilcoxon signed-rank test.

Ensemble	MAE after bias correction	MAE prior to bias correction	Std dev after correction
GFS 0000 UTC	1.67	1.99	0.74
GFS 1800 UTC	1.66	2.05	0.80
NAM 0000 UTC	1.68	1.91	0.67
NAM 1800 UTC	1.70	1.93	0.73
Deterministic forecast	1.70	1.77	—
Final OP ensemble	*1.52*	**1.67**	0.98

average wind speeds on the order of 10 m s^{-1}, would imply a change in errors of roughly 1% and thus a decrease in losses by $12 000 000 over the lifetime of the facility. The improvement would likely be larger if seasonally varying bias corrections were used, and if the improvements in speed forecasts are largest within the portion of the power curve where power generation is most sensitive to the wind speed.

5. Evaluation of ramp event forecasts

Among the biggest challenges facing the wind industry are sharp and sudden increases or decreases in the wind speed. Called ramp events, these periods have major effects on wind power production. In Figs. 7 and 8, we examined a yearly climatology of 2- and 4-h ramp-up and ramp-down events at the Pomeroy meteorological tower. We observed a strong bimodal peak cycle with maxima in December and June and a smooth decrease between these peaks with minima in February–March and August for both ramp-up and ramp-down events (Fig. 7). Ramp-up events showed a frequency of over 1.35 day^{-1} on average during December and June and a minimum of 0.9 day^{-1} on average during February and August. Ramp-down events showed a frequency of over 1.4 day^{-1} on average during December and June and a minimum of 0.8 day^{-1} on average during March and August. Regarding 4-h events versus 2-h events, the trends were very similar. However, the colder portion of the year had more longer-period ramp-ups than the warmer part, as there was a larger decrease in the 2-h event rates compared to 4-h events during this time. During the warmer part of the year, most ramp-ups happen quickly, whereas a larger fraction of winter ramp-ups take longer than 2 h to occur. This is likely due to increased convective activity over Iowa during the summer months, resulting in very quick ramp-up events. Such trends were not obvious for ramp-downs, however.

In addition, we examined the hourly climatology of ramp events (Fig. 8) and found that ramp-up events are most common during 0000–0300 UTC (1801–2100 LST), which is likely associated with the decoupling of the surface layer as the ground begins to cool, an event that has been used to explain the formation of the LLJ. Ramp-down events are most common during 1200–1500 UTC (0600–0900 LST), which is likely due to the coupling of the surface layer as the ground begins to warm, a time when LLJs are typically ending. The general diurnal trends do not differ as the ramp definition is changed from 4 to 2 h. The primary change, as expected, is a small, roughly 10%–20%, decrease in the total number of ramps when using the smaller time window. There is no consensus yet as to what time interval should define a ramp; however, it is encouraging that a change between 2 and 4 h does not lead to big differences in the results.

Because ramp events have a high level of impact on the wind energy industry, we examined in more detail the ability of the WRF to forecast these events. Please note that this comparison of ramp events with the WRF was done over a subset of the yearly climatology used previously, not the entire dataset. Table 12 shows the number of ramp-up, ramp-down, and total ramp events for both the day 1 (6–30 h after model start up) and day 2 (30–54 h after model start up) periods. All PBL schemes on day 2 and all PBL schemes except for the MYNN 2.5 scheme on day 1 forecasted a significantly lower number of ramp events than observed, according to the Wilcoxon signed-rank test. This suggests that the forecast models may be showing more gradual transitions during events, such that the wind speed changes do not meet the definition of a ramp. During days 1 and 2, the YSU scheme forecasted the fewest number of total ramp events, less than half of the number observed. This underprediction of the model was echoed in a study by Bradford et al. (2010), in which a privatized version of the 3-km WRF significantly underestimated the number of surface ramp events over an area of northern Texas, western Oklahoma, and southern Kansas.

It was initially assumed that most ramp events would be associated with either frontal passage or the presence of thunderstorms in this data subset, but these phenomena accounted for only 16% and 12%, respectively, of all 4-h ramps (Fig. 9). Although some events did occur during these weather phenomena, 28% of the events happened without an obvious trigger being present during 4-h

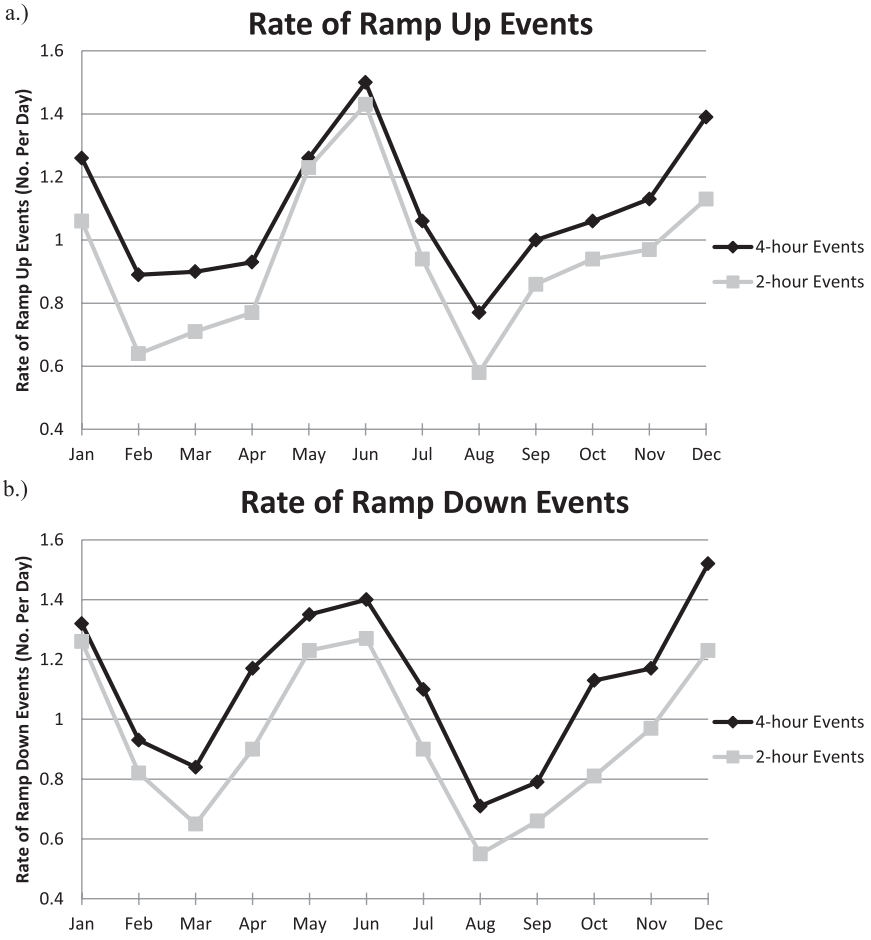

FIG. 7. Monthly climatology of (a) ramp-up and (b) ramp-down events per day over a 4-h (black) and 2-h (gray) window.

ramps. During 29% of all 4-h ramp events, an LLJ existed, and it is possible that mechanical mixing brought stronger winds down during short periods. In other events, the only weather condition noted that seemed as though it could play a role in rapid changes in wind speed was the presence of rather steep lapse rates near the surface, which could support propagating gravity waves of growing amplitude that become nonlinear, break, and create a high-wind episode at low levels. Fifteen percent of all ramp-up events occurred during the mid- or late morning when one might expect winds to increase quickly near the ground as the PBL grows, and a few ramp-down events happened toward evening when the collapse of the PBL might explain the decrease. But these events that appeared to be linked to diurnal changes in the PBL did not dominate the sample.

Halving the time window of the ramp criteria to 2 h resulted in a slightly higher fraction of ramp events being caused by frontal passages and thunderstorms, with each of these phenomena accounting for 17% of all ramps. Still, a large percentage of events (24%) happened without an obvious trigger being present, and 32% of the cases occurred while an LLJ was present. Ten percent of the events were likely caused by the growth of the PBL. Future work should examine in more detail possible causes of ramp events like these, but enhanced observational facilities might be necessary to explain the relatively large fraction of events for which no cause can be identified from the standard observational network.

Using the midpoint of each modeled ramp event, frequency of occurrence as a function of hour (within 3-h bins) was compared to observations for both ramp-up and ramp-down events (Fig. 10). Model ramp-up events occurred most frequently between 2200 UTC (1600 LST) and 0100 UTC (1900 LST) (late afternoon) in all schemes except YSU, while observed ramp-up events occurred most frequently around 0100 UTC (1900 LST). This sharp increase around 0100 UTC (1900 LST), as seen in the observed data for this period, is similar to the yearly climatology data. Again, this increase is likely associated

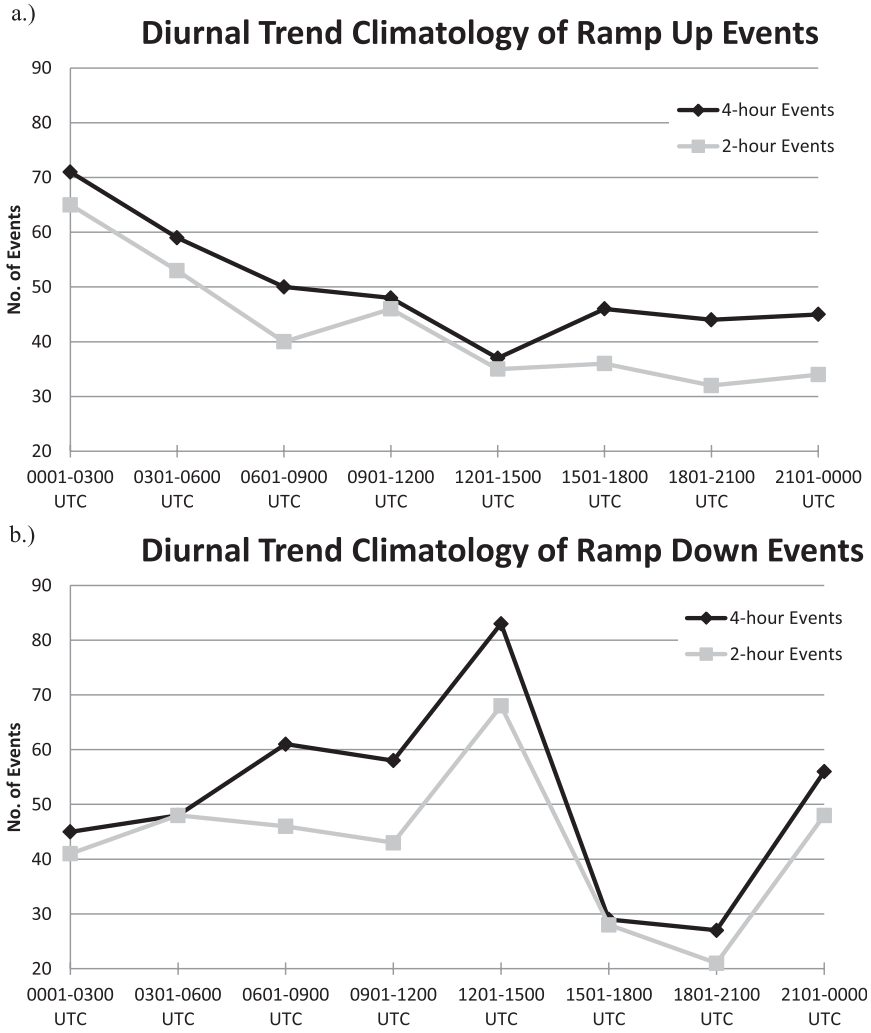

FIG. 8. Diurnal trend climatology of ramp-up (black) and ramp-down (gray) events (total number over a 1-yr period centered within 3-h bins) using (a) 4- and (b) 2-h definitions of ramps.

with the decoupling of the surface layer as the ground begins to cool, an event that has been used to explain the formation of the LLJ. It is of note that the development of an LLJ may be a layered substructure phenomenon, especially when a pronounced residual PBL persists above the LLJ. Much of emphasis so far in this paper has been put on the advantages of nonlocal schemes, but for this phenomenon, the YSU seems to have the most trouble, while the local mixing schemes predict the 0100 UTC maximum well. This is likely because layered substructures are more dependent upon local calculations. In this subsample of mostly fall cases, a secondary ramp-up peak, occurring around 1600 UTC (1000 LST), is more pronounced than in the yearly climatological data. This ramp-up event may be due to the growth of the boundary layer in the morning hours, which would be a period when higher-momentum air would begin mixing downward. During rare situations, if the hub height was located within the high-friction surface layer in the early morning, then the growth of the PBL should allow for potentially rapid increases in wind speed, assuming the LLJ peaks in intensity above turbine height. Future work should examine in more detail the frequency of occurrence of LLJs with peak speeds at or below hub height. However, a more likely explanation is that penetrating plumes are entraining higher-momentum air into the PBL (in non-LLJ conditions), and causing this ramp-up. Only the YSU scheme predicted this secondary maximum at this time of day, which suggests that its explicit entrainment is performing reasonably well. No other scheme indicated a secondary maximum during this mid- to late morning period. Thus, from a timing standpoint, the YSU scheme stands out as being substantially different from the other five schemes during ramp-up events.

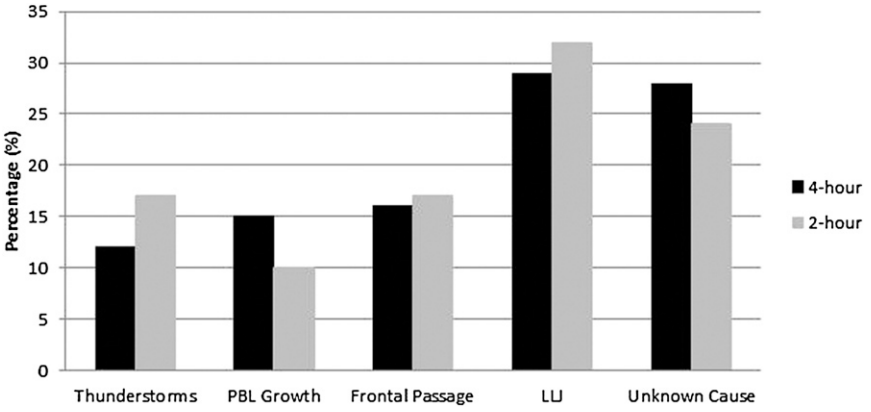

FIG. 9. Percentage of ramp events by most likely cause over a 4- (black) or a 2-h (gray) window.

For ramp-down events, a temporal trend in the observed data was less clear. Slight maxima in both the observations and in all of the PBL schemes except the YSU were observed around 0400 UTC (2200 LST) and 1300 UTC (0700 LST), although they were not as well defined as the ramp-up maxima. The sharp peak between 1200 and 1500 UTC (0600 and 0900 LST) in the yearly climatological data is still noticeable; however, it is less pronounced in the mostly fall dataset. Minima were observed around 0700 UTC (0100 LST) and 1900 UTC (1300 LST). The MYNN 2.5 and 3.0 PBL schemes captured the 1900 UTC (1300 LST) minimum, but none captured the one occurring at 0700 UTC (0100 LST). Once again, the YSU scheme's behavior was distinctly different from the others with its peak at 0100 UTC (1900 LST), a time when other schemes showed a distinct minimum.

To quantify timing error, MAE and bias were used to compare the different PBL schemes (Table 12). Note that MAE and bias values can only be determined for forecasted events, as there is no way to put a timing error on an event that was never forecasted. However, POD, FAR, and TS can be computed for both forecasted and missed model ramp events. As a result, a PBL scheme can have a small MAE and bias, but also a small POD. In all cases during this study, MAE was much larger than the bias, indicating that the PBL schemes were inconsistent with the timing of the ramp events. Ramp-up events had a higher MAE compared to ramp-down events in all PBL schemes, implying ramp-down events had better timing predictions than ramp-up events, although due to the subjective nature of defining the start of ramp events, caution must be used in interpreting these results.

Model error also was analyzed based on hits, misses, and false alarms. A hit was defined as a model ramp event occurring within ±6 h of an observed ramp event of the same type (observed ramp-up to modeled ramp-up). Most ramp-up hits, false alarms, and ramp-up events forecasted were associated with the MYNN 2.5 PBL scheme (Table 12). The high number of hits was due to the fact that this scheme forecasted the most events, and it was not associated with high model accuracy. For the ramp-down events, the QNSE scheme had the most forecasted ramp-down events, hits, and false alarms (tie), and, again, the high number of hits was due to the high number of events forecasted and was not associated with high model accuracy.

Therefore, to assess the ability of the various model runs, POD, FAR, and TS were calculated. Values of POD, FAR, and TS range from 0 to 1 with the more accurate models having a POD and TS near 1, and an FAR near zero. For all PBL schemes except YSU and Pleim, ramp-up events had higher POD scores, implying that models predict ramp-up events more accurately compared to ramp-down events. The MYNN 2.5 PBL scheme showed the best POD, detecting ramp-up events nearly 50% of the time. As expected, day 1 ramp events had higher PODs in all PBL schemes except the Pleim scheme, as forecast accuracy typically decreases with increasing lead time. Except for the YSU and Pleim schemes, a higher FAR was associated with ramp-down events compared to ramp-up events, implying models tend to forecast ramp-down events more often when observed ramp-down events are not present. The MYNN 2.5 PBL scheme showed the worst FAR, 0.50 or more on both days. Finally, in all schemes but the YSU and Pleim, the TS was higher for ramp-up events than for ramp-down events, confirming more accurate predictions of ramp-up events than ramp-down events. The scheme with the best detection (highest TS) for ramp-up events was the MYNN 2.5 PBL scheme, while the Pleim PBL scheme had the best detection for ramp-down events.

6. Summary and conclusions

Understanding the biases and strengths of different PBL schemes will help to improve wind speed forecasts

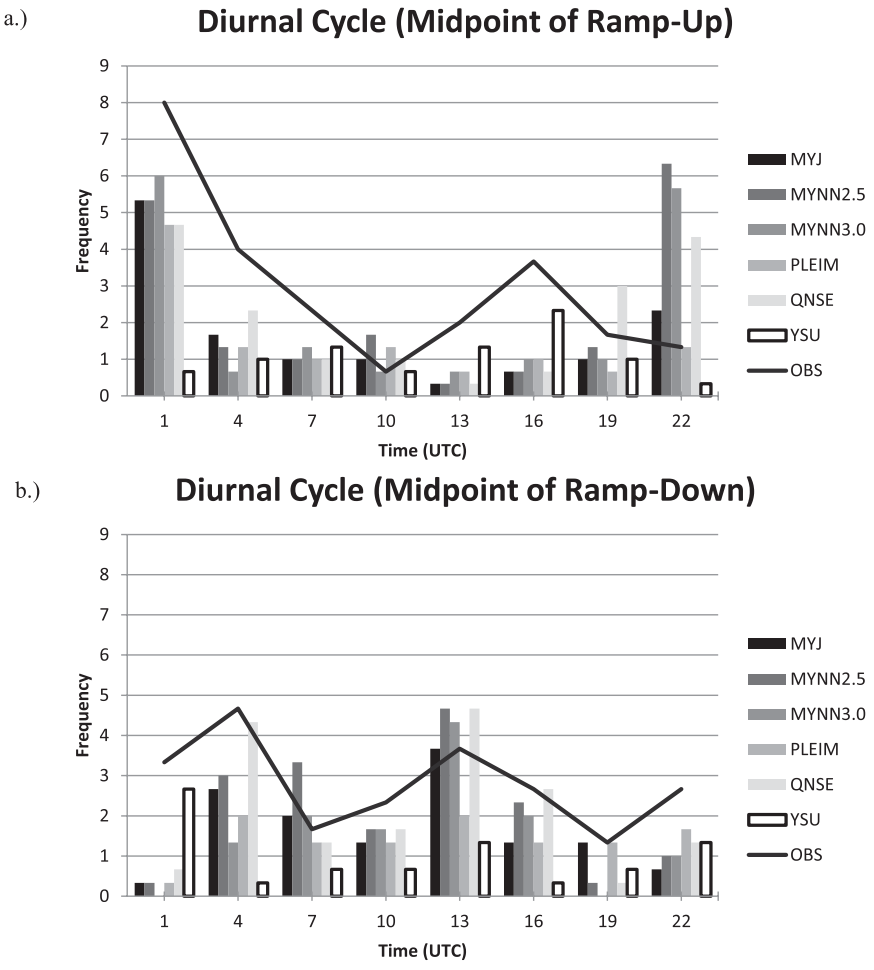

FIG. 10. Number of (a) ramp-up and (b) ramp-down events centered within 3-h bins using the midpoint time of the ramp events. Model runs are indicated with grayscale bar on right, with black line representing observed events.

at 80 m. In an examination of ensemble designs, it was found that perturbations of the GFS ILBCs resulted in larger model spread than that achieved with the use of six PBL schemes; however, the MAE of the ensemble mean was higher with the GFS perturbations. Simulations using the GFS ILBCs also showed lower MAE than those using the NAM. Finally, ensembles using different time initializations gave larger spread and lower MAE than the GFS perturbations tested.

The first postprocessing technique examined, training the model based on day 1 results, was found to yield a forecast with high MAE as conditions apparently change too much from day to day. The second technique tested, the neighborhood approach, increased the accuracy of the models, although not significantly. The postprocessing technique that was most successful was bias correction. Many different bias corrections were tested; however, the wind speed bias correction yielded the best results. From these results, a six-member operational ensemble was developed that significantly outperformed other ensembles tested. Of the six members, the nonlocal mixing schemes of the Pleim and YSU formed five out of the six members, indicating, at least for this study, that nonlocal schemes did better than local schemes when predicting 80-m wind speed.

Many impediments preclude accurate forecasts of wind conditions at 80 m. We know at the surface winds decrease at night due to the decoupling of the surface layer, and increase during the day as the boundary layer grows and higher momentum air from above mixes down. However, with very few observations at 80 m, we have yet to develop a robust method for forecasting the time evolution of wind speed between the middle PBL and the surface. As a result, unforecasted ramp events, sharp increases or decreases in wind speed over a small time period, reduce the reliability of wind as a source of power.

TABLE 12. Number of ramp events during day 1 (6–30 h after model start up) and day 2 (30–54 h after model start up) and model error associated with ramp events for each PBL scheme. POD, FAR, and TS values were calculated. The bias and MAE show the timing error associated with each PBL scheme. A hit means the model correctly predicted the ramp event within ±6 h. Boldface values indicate best POD, FAR, and TS scores while italicized values indicate worst POD, FAR, and TS scores.

PBL scheme	Ramp type	Obs total events	Model total events	Hits	False Alarm	Miss	MAE (h)	Bias (h)	POD	FAR	TS
MYJ	Up (day 1)	35	23	17	6	18	3.47	−1.24	0.49	**0.26**	0.41
	Up (day 2)	37	17	13	4	24	1.85	−1.23	0.35	**0.32**	0.32
	Down (day 1)	31	20	8	12	23	1.88	0.63	0.26	0.60	0.19
	Down (day 2)	35	19	12	7	23	1.42	−0.42	0.34	0.37	0.29
MYNN 2.5	Up (day 1)	35	29	19	10	16	2.68	−1.74	**0.54**	0.34	**0.42**
	Up (day 2)	37	25	15	10	22	2.33	−1.20	**0.41**	0.40	**0.32**
	Down (day 1)	31	28	11	17	20	1.64	−0.73	0.35	*0.61*	0.23
	Down (day 2)	35	22	11	11	24	1.55	−0.27	0.31	*0.50*	0.24
MYNN 3.0	Up (day 1)	35	27	17	10	18	2.88	−1.71	0.49	0.37	0.38
	Up (day 2)	37	24	16	8	21	2.75	−1.13	0.43	0.33	0.36
	Down (day 1)	31	21	9	12	22	1.89	−0.56	*0.29*	0.57	*0.21*
	Down (day 2)	35	16	8	8	27	1.50	0.25	*0.23*	0.50	*0.19*
Pleim	Up (day 1)	35	19	10	9	25	3.10	−1.30	0.29	0.47	0.23
	Up (day 2)	37	17	12	5	25	2.33	−1.83	0.32	0.29	0.29
	Down (day 1)	31	14	9	5	22	2.22	0.44	0.29	0.36	**0.25**
	Down (day 2)	35	20	12	8	23	2.00	0.50	0.34	0.40	**0.28**
QNSE	Up (day 1)	35	26	18	8	17	3.56	−2.56	0.51	0.31	0.42
	Up (day 2)	37	26	15	11	22	1.73	−1.20	0.41	0.42	0.31
	Down (day 1)	31	28	11	17	20	1.27	−1.00	**0.35**	0.61	0.23
	Down (day 2)	35	23	12	11	23	1.33	−0.22	**0.34**	0.48	0.26
YSU	Up (day 1)	35	16	8	8	27	3.25	−0.25	*0.23*	*0.50*	*0.19*
	Up (day 2)	37	11	8	3	29	2.50	0.25	*0.22*	*0.27*	*0.20*
	Down (day 1)	31	13	9	4	22	1.33	−0.22	0.29	**0.31**	0.26
	Down (day 2)	35	11	9	2	26	1.33	−0.89	0.26	**0.18**	0.24

For ramp events at 80 m, we found that all six PBL schemes tested underestimated the number of ramp-up and ramp-down events. Regarding frequency of occurrence, modeled ramp-up events occurred most often between 2200 UTC (1600 LST) and 0100 UTC (1900 LST), which closely matched observed ramp-up events [most frequent around 0100 UTC (1900 LST)]. The sharp increase in observed ramp-up events around 0100 UTC (1900 LST) may be associated with the decoupling of the surface layer as the ground began to cool, leading to the formation of the LLJ. In all ramp events, MAE was larger than the bias, indicating that the PBL schemes were inconsistent with the timing of the ramp events. For ramp-up events, all local mixing schemes had higher POD, lower FAR, and higher TS than the nonlocal schemes, implying that the local mixing schemes exhibit greater ability in simulating internal structures like LLJ development. This prompts us to draw two conclusions: first, vastly more observations of wind and temperature in the lowest 500 m of the PBL are needed under all conditions to establish a climatology for this region and, second, guided by these observations we need to reexamine the representations (local, nonlocal, turbulence order) of turbulent processes of PBL schemes used to represent mixing processes in this layer.

Acknowledgments. We thank MidAmerican Energy Company for providing the 80-m observations. Partial funding was supplied by NSF Grant BCS0618823, DOE Grant 13-450-141201, Ames Laboratory Project 290-25-09-02-0031, and EPRC Grant 400-60-12. The helpful comments of three anonymous reviewers were appreciated.

APPENDIX

Description of PBL Schemes

PBL schemes were developed to help resolve the turbulent fluxes of heat, moisture, and momentum in the boundary layer. However, due to the complex nature of turbulence, closure has remained a problem. Two solutions to the problem of closure, local and nonlocal, will be discussed below. The first type, local closure, estimates unknown fluxes using known values and/or gradients at the same point. The second type, nonlocal closure, estimates unknown fluxes using known values and/or gradients at many points in space (Stull 1988, 197–242; Bélair et al. 1999). Of the PBL schemes tested, the ACM2 and YSU schemes are nonlocal while the MYJ, QNSE, and MYNN 2.5 methods are local closure schemes. Although

the calculations are computed locally with the MYNN 3.0 scheme, the higher-order terms (i.e., temperature variance) have a countergradient term, which helps to parameterize the effects of nonlocal mixing (Stensrud 2007, 169–171; Nakanishi and Niino 2009). A brief description of the six PBL schemes used in this study follows. Further details can be found in Janjić (1990, 1994) for MYJ, Hong et al. (2006) for YSU, Pleim (2007a,b) for ACM2, Sukoriansky et al. (2005) for QNSE, and Nakanishi and Niino (2009) for MYNN.

The MYJ PBL scheme is one of four different local closure schemes evaluated in this study. The MYJ PBL scheme is a local turbulent kinetic energy (TKE), 1.5-order (2.5-level) closure scheme. Being a 1.5-order closure scheme, it requires one additional prognostic equation to solve for the turbulent quantities (Janjić 1990, 1994; Shin and Hong 2011; Hu et al. 2010). The MYNN 2.5 and 3.0 PBL schemes are higher-level schemes that were based on the MYJ approach. The MYNN 2.5 scheme is a local TKE, 1.5-order (2.5-level) closure scheme while the MYNN 3.0 is a local TKE, 2.0-order (3.0-level) closure scheme. Both the MYJ and MYNN schemes apply the local mixing from the lowest to highest vertical level. The major difference between the MYJ and MYNN 2.5 and 3.0 schemes is the TKE equation and, more specifically, the master mixing length (l_m). The TKE equation is defined as

$$\frac{d(q^2/2)}{dt} - \frac{\partial}{\partial z}\left[l_m q S_q \frac{\partial}{\partial z}\left(\frac{q^2}{2}\right)\right] = P_s + P_b + \varepsilon,$$

where the first term is the total derivative of q, which is 2 times the TKE; the second term is the vertical redistribution of q; P_s is the production of q by shear; P_b is the production of q by buoyancy; and ε is the dissipation term. For the MYJ scheme, the master mixing length is defined as

$$l_m = l_o \frac{kz}{kz + l_o},$$

where l_o is dependent on height and k is the von Kármán constant. The master mixing length for the MYNN PBL schemes is a function of three independent length scales:

$$\frac{1}{l_m} = \frac{1}{l_s} + \frac{1}{l_t} + \frac{1}{l_b},$$

where l_s is the surface layer length, l_t is the turbulent layer length, and l_b is the buoyancy length (Olson and Brown 2009; Nakanishi and Niino 2009).

The QNSE scheme is a local TKE, 1.5-order (2.5-level) closure scheme that is similar to the MYJ scheme during neutral and unstable conditions. The QNSE scheme differs from the MYJ scheme during stable conditions, when spectral theory is used to develop eddy diffusivity profiles. This results in waves and turbulent eddies being treated as one entity. Like the MYJ and MYNN schemes, the QNSE scheme applies local mixing from the lowest to highest vertical levels (Sukoriansky et al. 2005; Shin and Hong 2011).

The last two PBL schemes investigated in this study were the YSU and ACM2. These schemes are both first-order (requiring no additional prognostic equations), nonlocal approaches. The ACM2 scheme is a combination of a simple transilient model (original Blackadar scheme) and an eddy diffusion model. The ACM2 scheme is able to switch between stable conditions (eddy diffusion) and unstable conditions (local and nonlocal transport). During stable or neutral conditions, the scheme uses local closure instead of nonlocal transport (Hu et al. 2010; Pleim 2007a,b; Shin and Hong 2011). On the other hand, the YSU scheme is a bulk scheme that expresses nonlocal mixing by convective large eddies. Nonlocal mixing is achieved by adding a nonlocal gradient adjustment term (countergradient term) to the local gradient. At the top of the PBL, the YSU scheme uses explicit treatment of the entrainment layer, which is proportional to the surface layer flux (Hong et al. 2006; Shin and Hong 2011; Hu et al. 2010).

REFERENCES

Ayotte, K. W., R. J. Davy, and P. A. Coppin, 2001: A simple temporal and spatial analysis of flow in complex terrain in the context of wind energy modeling. *Bound.-Layer Meteor.,* **98,** 275–295.

Bélair, S., J. Mailhot, J. W. Strapp, and J. I. MacPherson, 1999: An examination of local versus nonlocal aspects of a TKE-based boundary layer scheme in clear convective conditions. *J. Appl. Meteor.,* **38,** 1499–1518.

Bradford, K. T., R. L. Carpenter, and B. Shaw, 2010: Forecasting Southern Plains wind ramp events using the WRF model at 3-km. Preprints, *Ninth Annual Student Conf.,* Atlanta, GA, Amer. Meteor. Soc., S30. [Available online at http://ams.confex.com/ams/pdfpapers/166661.pdf.]

Briggs, W. M., and D. Ruppert, 2004: Assessing the skill of yes/no forecasts for Markov observations. Preprints, *17th Conf. on Probability and Statistics in the Atmospheric Sciences,* Seattle, WA, Amer. Meteor. Soc., 2.2. [Available online at http://ams.confex.com/ams/pdfpapers/68436.pdf.]

Dalcher, A., E. Kalnay, and R. N. Hoffman, 1988: Medium range lagged average forecasts. *Mon. Wea. Rev.,* **116,** 402–416.

Department of Energy, 2008: 20% wind energy by 2030. Energy Efficiency and Renewable Energy Rep. DOE/GO-102008-2567, 248 pp.

——, 2010: Electric power annual 2009. U.S. Energy Information Administration Rep. DOE/EIA-0348, 108 pp.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.,* **46,** 3077–3107.

Ebert, E. E., 2009: Neighborhood verification: A strategy for rewarding close forecasts. *Wea. Forecasting,* **24,** 1498–1510.

Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, and V. Koren, 2003: Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta Model. *J. Geophys. Res.,* **108,** 8851, doi:10.1029/2002JD003296.

Ferrier, B. S., Y. Jin, T. Black, E. Rogers, and G. DiMego, 2002: Implementation of a new grid-scale cloud and precipitation scheme in NCEP Eta model. Preprints, *15th Conf. on Numerical Weather Prediction,* San Antonio, TX, Amer. Meteor. Soc., 280–283.

Francis, N., 2008: Predicting sudden changes in wind power generation. *North Amer. Windpower,* **5,** 58–60.

Freedman, J., M. Markus, and R. Penc, 2008: Analysis of west Texas wind plant ramp-up and ramp-down events. AWS Truewind LLC Rep. prepared for GE Wind, Albany, NY, 250–278.

General Electric Company, 2005: 1.5 MW series wind turbine. GE Energy Rep. A4_BR APPENDIX A 1A- GE OVERVIEW, 4 pp.

Greaves, B., J. Collins, J. Parkes, and A. Tindal, 2009: Temporal forecast uncertainty for ramp events. *Wind Eng.,* **33,** 309–320.

Hall, T. J., R. N. Thessin, G. J. Bloy, and C. N. Mutchler, 2010: Analog sky condition forecasting based on a *k*-nn algorithm. *Wea. Forecasting,* **25,** 1463–1478.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: Two transplant case studies. *Quart. J. Roy. Meteor. Soc.,* **125,** 2487–2516.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.,* **124,** 2322–2339.

——, Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with explicit treatment of entrainment processes. *Mon. Wea. Rev.,* **134,** 2318–2341.

Houtekamer, P. L., 1993: Global and local skill forecasts. *Mon. Wea. Rev.,* **121,** 1834–1846.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF model. *J. Appl. Meteor. Climatol.,* **49,** 1831–1844.

Huang, H., and Z. S. Chalabi, 1996: Use of time-series analysis to model and forecast wind speed. *J. Wind Eng. Ind. Aerodyn.,* **56,** 311–322.

Janjić, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.,* **118,** 1429–1443.

——, 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous layer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.,* **43,** 170–181.

Lu, C., H. Yuan, B. Schwartz, and S. Benjamin, 2007: Short-range forecast using time-lagged ensembles. *Wea. Forecasting,* **22,** 580–595.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.,* **102** (D14), 16 663–16 682.

Nakanishi, M., and H. Niino, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan,* **87,** 905–909.

Olson, J. B., and J. M. Brown, 2009: A comparison of two Mellor–Yamada-based PBL schemes in simulating a hybrid barrier jet. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., JP1.13. [Available online at http://ams.confex.com/ams/pdfpapers/154321.pdf.]

Pleim, J. E., 2007a: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.,* **46,** 1383–1395.

——, 2007b: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *J. Appl. Meteor. Climatol.,* **46,** 1396–1409.

——, and A. Xiu, 1995: Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models. *J. Appl. Meteor.,* **34,** 16–32.

Schreck, S., J. Lundquist, and W. Shaw, 2008: U.S. Department of Energy workshop report: Research needs for wind resource characterization. NREL Rep. TP-500-43521, 107 pp.

Shin, H. H., and S. Hong, 2011: Intercomparison of planetary boundary-layer parametrizations in the WRF model for a single day from CASES-99. *Bound.-Layer Meteor.,* **139,** 261–281.

Stensrud, D. J., 2007: *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge University Press, 459 pp.

——, J.-W. Bao, and T. T. Warner, 2000: Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Wea. Rev.,* **128,** 2077–2107.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Kluwer Academic, 666 pp.

Sukoriansky, S., B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stable stratified turbulence to the atmospheric boundary layer over sea ice. *Bound.-Layer Meteor.,* **117,** 231–257.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.,* **12,** 257–268.

Walser, A., D. Lüthi, and C. Schär, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.,* **132,** 560–577.

Wei, M., Z. Toth, R. Wobus, Y. Zhu, C. H. Bishop, and X. Wang, 2006: Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP. *Tellus,* **58A,** 28–44.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.,* **126,** 3292–3302.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.

Wood, N., 2000: Wind flow over complex terrain: A historical perspective and the prospect for large-eddy modeling. *Bound.-Layer Meteor.,* **96,** 11–32.

Xiu, A., and J. E. Pleim, 2001: Development of a land surface model. Part I: Application in a mesoscale meteorological model. *J. Appl. Meteor.,* **40,** 192–209.

Zack, J. W., 2007: Optimization of wind power production forecast performance during critical periods for grid management. *Windpower 2007,* Los Angeles, CA, Amer. Wind Energy Association, 12 pp.

Zhang, D.-L., and W.-Z. Zheng, 2004: Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations. *J. Appl. Meteor.,* **43,** 157–169.