Wind ramp events at an Iowa wind farm: A climatology and evaluation of WRF ensemble forecast skill

William A. Gallus, Jr.[1]
Adam J. Deppe[1]

Department of Geological and Atmospheric Sciences, Iowa State University[1]

1. Introduction

Wind, unlike other sources of energy, varies substantially over both space and time resulting in production rates that fluctuate more strongly than other traditional fossil fuel sources of energy generation. 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 (AWEA 2009). However, the expanding wind energy field will require better prediction of winds. Perhaps the most difficult challenge in forecasting winds is the accurate prediction of ramp events, which are defined as rapid changes in wind speed that lead to extreme changes in wind power output. Because power production sharply increases between the cut-in speed and the rated wind speed, ramp events in this area are extremely costly to energy companies (Francis 2008). Accurate forecasts of these events could greatly benefit the industry, but because these events occur over short temporal scales and may also be of limited extent spatially, forecasting them can be difficult.

In this study, the ability of the Weather Research and Forecasting (WRF) model to accurately reproduce ramp events at hub height (80m) was evaluated by comparing WRF simulations using six different planetary boundary layer (PBL) schemes to observations of 80m wind speed gathered at the Pomeroy, Iowa wind farm site.

Corresponding author address: William A. Gallus, Jr., ISU, 3010 Agronomy Hall, Ames, IA 50011. Email: wgallus@iastate.edu.

2. Data and Methodology

The WRF model with a 10-km horizontal resolution was run for 54 hours starting at 00 UTC to simulate ramp events, although validation was from 06-54 hours due to model start up and to cover the day 1 and day 2 periods of most interest to local wind energy companies. The following PBL schemes were evaluated: the Yonsei University scheme (YSU), the Mellor-Yamada-Janjic scheme (MYJ), the quasi-normal scale elimination PBL scheme (QNSE), the Mellor-Yamada Nakanishi and Niino level 2.5 PBL scheme (MYNN 2.5), the Mellor-Yamada Nakanishi and Niino level 3.0 PBL scheme (MYNN 3.0), and the Pleim PBL scheme (also called the asymmetric convective model [ACM2]). The model configurations above were run using the Global Forecast System (GFS) for initial and lateral boundary conditions. Sixty cases spanning 120 days from June 2008 through June 2009 were validated using hourly wind speed measurements at 80m from a meteorological tower at the Pomeroy wind farm in northwestern Iowa. These wind observations were put through an extensive quality control, and these cases were from the subset of days when reliable data existed. These wind data contained wind information every 10 minutes, and observed ramps were determined using both the 10 minute 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 one hour.

An event was considered to be a ramp event if the change in wind power was 50% or more of total capacity in four hours 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 within the 6-12 m/s window (where power production varies greatly) in four hours or less was considered a ramp. If a ramp occurred, it was classified into two categories: a ramp-up event (increase in wind speed within four hours) or a ramp-down event (a decrease in wind speed within four hours). A similar concept was used by Freedman et al. (2008) in a west Texas study that found that between 2005 and 2006, about 60% of the ramp events observed were ramp-up events while only 40% were ramp-down events. For the period of study at the Pomeroy wind farm, 51.8% of all ramps were ramp-up events while 48.2% were ramp-down events. When 10 minute data were used for classification, the number of ramp events more than doubled, but the distribution remained nearly 50% for each type.

3. Results

Tables 1 and 2 showed the number of ramp-up, ramp-down, and total ramp events for Day 1 (6-30 hours after model start up) and Day 2 (30-54 hours after model start up). All PBL schemes on both Days 1 and 2 forecasted a lower number of ramp events than observed, suggesting the model runs might be showing a more gradual transition during events than was actually occurring, thus causing a wind speed change to not meet the criteria for definition as a ramp. During Days 1 and 2, the YSU scheme had the fewest number of total ramp events, less than half of the number observed. This under-prediction of the model was echoed in a study by Bradford et al. 2010, in which a 3km WRF model significantly underestimated the number of surface ramp events over an area of northern Texas, western Oklahoma, and southern Kansas. The MYNN 2.5 scheme had the most ramp events forecasted during Day 1 while the QNSE scheme had the most ramp events during Day 2. Both

PBL Scheme	MYJ	MYNN 2.5	MYNN 3.0	Pleim	QNSE	YSU	Obs
Ramp-up	23	29	27	19	26	16	35
Ramp-down	23	28	21	14	28	13	31
Total Ramp Events	46	57	48	33	54	29	66

Table 1: Number of ramp events during Day 1 (06-30 hours after model start up).

PBL Scheme	MYJ	MYNN 2.5	MYNN 3.0	Pleim	QNSE	YSU	Obs
Ramp-up	17	25	24	17	26	11	37
Ramp-down	19	22	16	20	23	11	35
Total Ramp Events	36	47	40	37	49	22	72

Table 2: Number of ramp events during Day 2 (30-54 hours after model start up).

Tables 1 and 2 show that the number of observed ramp-up and ramp-down events is similar on Days 1 and 2. This was not the case for each individual PBL scheme as a lower number of ramp events occurred on Day 2 compared to Day 1.

An analysis was performed to try to identify causes for all of the observed ramp events. It was initially assumed that most would be associated with either frontal passage or the presence of thunderstorms. Although some events did occur during these weather phenomena, most events happened without an obvious trigger being present. In some cases, a low-level jet existed, and it is possible that mechanical mixing brought stronger winds down during short periods. In other events, the only item noted was the presence of rather steep lapse rates near the surface, which could facilitate enhanced mixing of higher winds toward the surface at some times. Some ramp up events did occur during the mid or late morning when one might expect wind 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.

Forecasting the correct amplitude of ramp events will help managers prevent grid overloads and blackouts due to low power resources. From the 60 cases tested, amplitude was over-predicted by all six PBL schemes for ramp-up events during Day 1 and 2 (Table 3). This result suggested that the PBL schemes are overly aggressive in increasing the wind speed associated with ramp-up events. Ramp-up events were also predicted to have larger amplitudes than ramp-down events in all PBL schemes; however, no difference in amplitude between ramp-up and down events was seen in the observed data. One possible cause for the difference between model simulations and observations could be the fact that LLJ events may make up a high portion of the ramp-up events. Carter et al. (2011) found that the height of the peak wind speed in LLJs was predicted too low by the different PBL schemes. As a result, the models may be mixing higher momentum air downward too strongly, possibly resulting in an over-prediction of the ramp-up amplitudes. No bias or trend was associated with ramp-down events as results are mixed with some schemes over-predicting and others under-predicting amplitude, although the ramp-down events showed a lower Mean Absolute Error (MAE) compared to the ramp-up events.

PBL Scheme	MYJ (m/s)	MYNN 2.5 (m/s)	MYNN 3.0 (m/s)	Pleim (m/s)	QNSE (m/s)	YSU (m/s)	Obs (m/s)
Ramp-up (Day 1)	4.50	4.62	4.75	4.85	4.60	4.67	4.53
Ramp-up (Day 2)	4.54	5.16	5.2	4.56	4.69	4.73	4.01
Ramp-down (Day 1)	3.74	4.62	4.20	4.60	4.31	4.17	4.34
Ramp-down (Day 2)	3.83	4.28	4.46	4.27	4.59	4.43	4.21

Table 3: Average amplitude of ramp events divided into ramp-up/down events on Day 1 and Day 2.

The duration of a ramp event is important to energy companies because the longer such an unpredicted ramp event lasts, the more money it costs (Francis 2008). The duration of modeled ramp-up and ramp-down events in our sample is much longer than observed as most PBL schemes showed ramp events lasting 1 to 1.5 hours longer (Table 4). We believe one possible reason is due to the lack of rigorous mixing in the different PBL schemes, causing the longer duration of ramp events. Longer model ramp duration was seen during Day 2 in all PBL schemes for all ramps, except the YSU for ramp-down events. Little difference is noted between the durations of ramp-up and ramp-down events, matching observations.

Understanding when a ramp-down or ramp-up event is likely to occur would greatly improve forecast prediction. Using a three hour average and the midpoint of the ramp event, ramp-up and ramp-down diurnal cycles were created (Figs. 1 and 2). Model ramp-up events occurred most frequently between 22Z and 01Z in all schemes except YSU, while observed ramp-up events occurred most frequently around 01Z. We believe that this sharp increase around 01Z in the observed data is 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 Low

PBL Scheme	MYJ (hr)	MYNN 2.5 (hr)	MYNN 3.0 (hr)	Pleim (hr)	QNSE (hr)	YSU (hr)	Obs (hr)
Ramp-up (Day 1)	3.30	3.14	3.48	3.05	3.46	3.31	2.34
Ramp-up (Day 2)	3.76	3.76	3.83	3.59	3.69	3.91	2.08
Ramp-down (Day 1)	3.13	3.71	3.76	3.36	3.39	3.62	2.32
Ramp-down (Day 2)	3.74	3.73	3.88	3.70	3.61	3.27	2.31

Table 4: Average duration of ramp-up/down events on Day 1 and Day 2.

Level Jet (LLJ), which is captured fairly well by the different PBL schemes, although initiation of the LLJ by the models is a couple of hours early (Carter et al. 2011). A secondary peak, occurring around 16Z can also be seen in the observed data. This ramp-up event is likely due to the growth of the boundary layer in the morning hours, which would be a period when higher momentum air might begin being mixed downward. Only the YSU scheme showed a peak at this time of day. No other scheme indicated a secondary maximum during this mid-late morning period. Thus, for timing of ramp-up events, the YSU scheme stands out as being dramatically different than the other five schemes. For ramp-down events, the amplitude of trends was less than the ramp-up events, with a hint in both observations and some simulations of maxima around 04 UTC and 13 UTC. Minima were observed around 07 UTC and 19 UTC. Some models did capture the 19 UTC minima, but none captured the one at 07 UTC. Once again, the YSU scheme behaved noticeably different from the others with its peak at 01 UTC, a time when other schemes showed a distinct minimum.

To quantify timing error, mean absolute error (MAE) and bias were used to compare the different PBL schemes to the observed data (Table 5). Bias values near zero with MAE values near zero indicate high model skill. In all cases, 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 prediction than ramp-up events. Again, this might be explained by the high frequency of LLJs in our sample that we believe contribute to ramp-up events. Carter et al. (2011) found that LLJs were predicted too early by the PBL schemes, which could be one reason why the ramp-up event time prediction is not as good as the ramp-downs.

Model error also was analyzed based on hits, misses, and false alarms. A hit was defined as a model ramp event occurring within +/- 6 hours of an observed ramp event of the same type (observed ramp-up to model ramp-up). The most ramp-up hits, false alarms, and total number of ramp-up events forecasted was associated with the MYNN 2.5 PBL scheme. The high number of hits was due to the fact that this scheme forecasted the most events, but the skill was not particularly high as many of the forecasted events were misses. 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 skill.

To further understand the skill of the various model runs, Probability of Detection (POD) (1), False Alarm Rate (FAR) (2), and Threat Score (TS) (3) were calculated using the following equations:

$$POD = \frac{Total\ number\ of\ correct\ event\ forecasts\ (Hits)}{Total\ number\ of\ events\ observed} \qquad \textbf{(1)}$$

$$FAR = \frac{Total\ number\ of\ false\ alarms}{Total\ number\ of\ events\ forecasted} \qquad \textbf{(2)}$$

$$TS = \frac{Total\ number\ of\ correct\ event\ forecasts\ (Hits)}{Total\ number\ of\ events\ forecasted + Number\ of\ misses} \qquad \textbf{(3)}$$

Values of POD, FAR, and TS range from 0 to 1 with high model skill having a POD and TS near 1, and FAR near zero. In all PBL schemes except YSU and Pleim, ramp-up events had higher POD scores, implying that models exhibit more skill in the detection of ramp-up events compared to ramp-down events. The MYNN 2.5 PBL scheme showed the best POD skill, detecting ramp-up events nearly 50% of the time. As expected, Day 1 ramp events had higher POD scores in all PBL schemes except the Pleim scheme, as forecast accuracy typically decreases with increasing lead time. Except in the YSU and Pleim scheme, a higher FAR score 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, .50 or more on both days. Finally, in all schemes but the YSU and Pleim, the TS was higher for ramp-up events than ramp-down events, confirming better model skill in detecting ramp-up events than ramp-down events. The scheme with the best detection skill (highest TS) for ramp-up events is the MYNN 2.5, with the Pleim scheme having the best detection skill for ramp-down events.

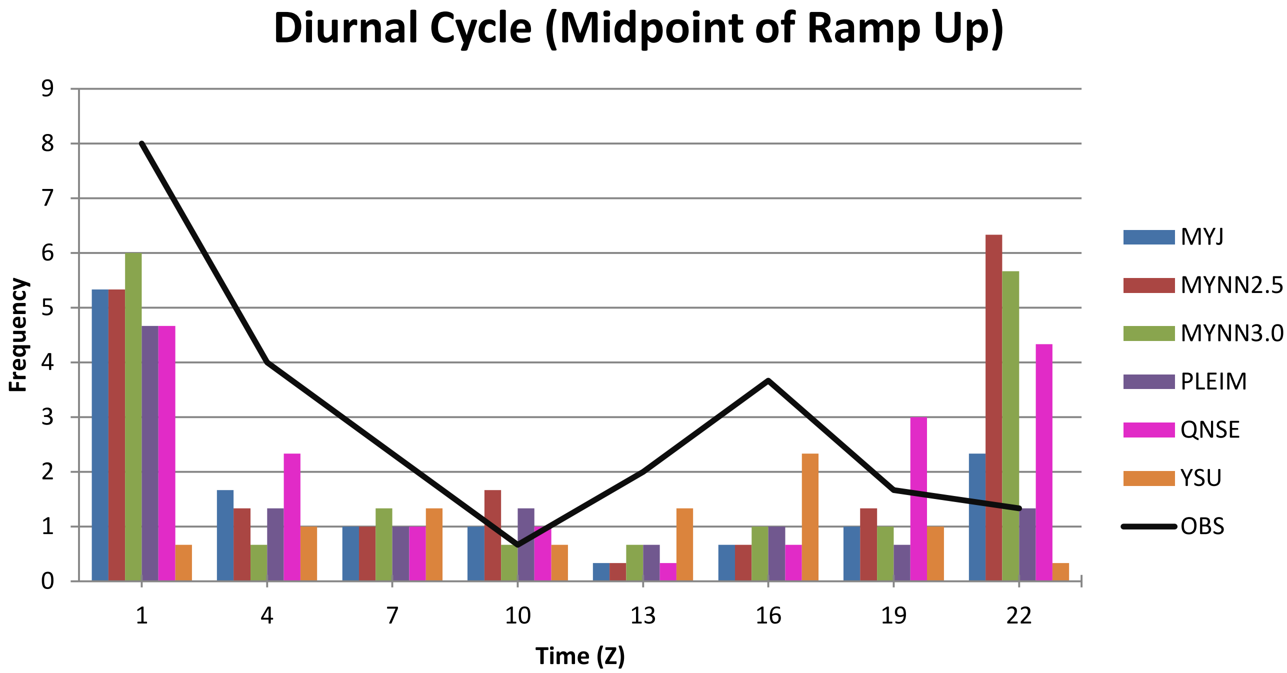

Fig. 1: Three hour averaged diurnal cycle of ramp-up events using the midpoint of the ramp event.

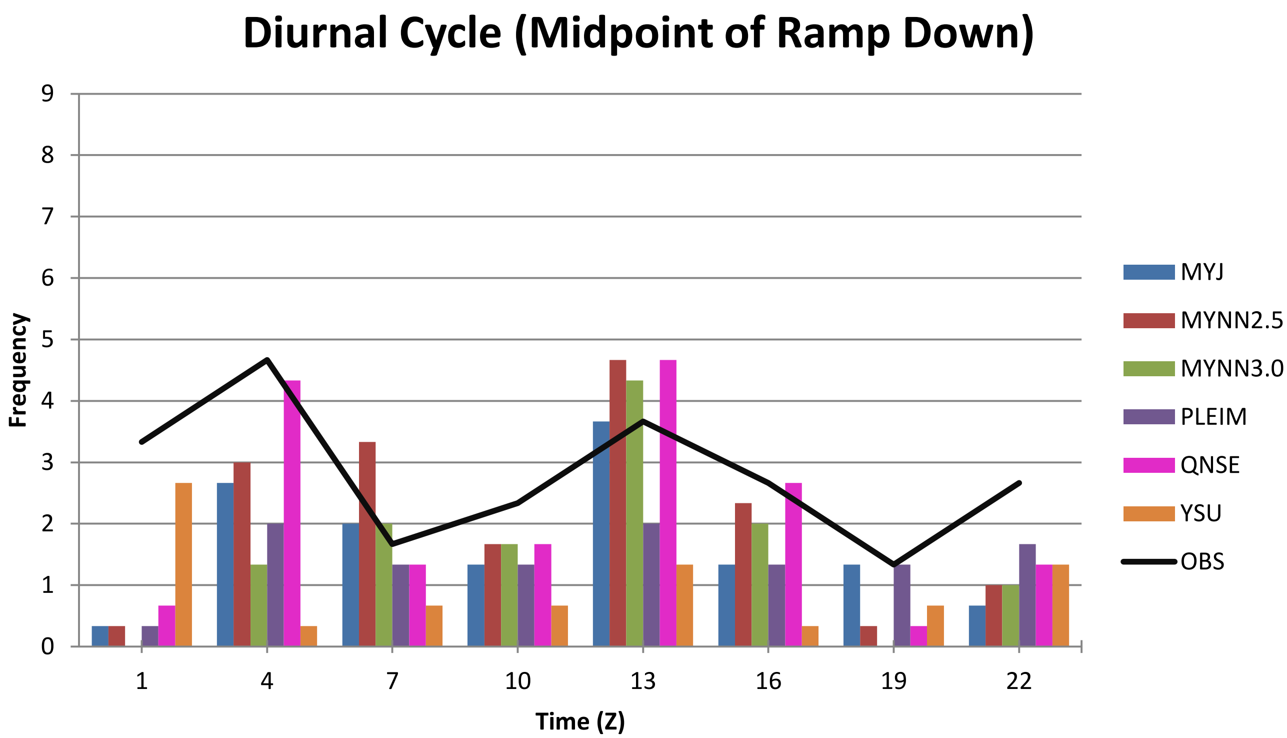

Fig. 2: Three hour averaged diurnal cycle of ramp-down events using the midpoint of the ramp event.

4. Summary and Conclusions

Understanding the biases and strengths of different PBL schemes should help to improve ramp event forecasts. From this study, we discovered that all six PBL schemes tested in the WRF model underestimated the number of ramp-up and ramp-down events compared to observations. Model ramp-up events had higher amplitudes than ramp-down events for all six PBL schemes, although no difference existed between the amplitude of observed ramp-up and ramp-down events. Larger model error was present in the amplitude of ramp-up events compared to ramp-down events. The duration of modeled ramp-up and ramp-down events was much longer than observed events, with most PBL schemes having ramp events lasting 1 to 1.5 hours longer than the observed ones. Regarding frequency, model ramp-up events occurred most often between 22Z and 01Z, while observed ramp-up events occurred most frequently around 01Z. In all cases, MAE was larger than the bias, indicating that the PBL schemes were inconsistent with the timing of the ramp events. In all PBL schemes except the YSU and Pleim, ramp-up events had higher POD, lower FAR, and higher TS, implying that models exhibit more skill in the detection of ramp-up events than ramp-down events. In conclusion, it appears forecasts are more skillful in detecting ramp-up events but the amplitude is better predicted in ramp-down events, and overall, similar to previous studies, there is much room for improvement in ramp forecasts from the WRF model running at this grid spacing.

5. Acknowledgments

We would like to thank MidAmerican Energy Company for providing the 80m observations. Partial funding was supplied by NSF grant BCS0618823, DOE grant #13-450-141201, Ames Laboratory Project 290-25-09-02-0031, and ERPC grant #400-60-12.

PBL Scheme	Ramp Type	Obs. Total Events	Model Total Events	Hits	False Alarm	Miss	MAE (hr)	Bias (hr)	POD	FAR	Threat Score
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

Table 5: Model error associated with ramp events for each PBL scheme. Probability of Detection (POD), False Alarm Rate (FAR) and Threat Score were calculated. The Bias and Mean Absolute Error (MAE) show the timing error associated with each PBL scheme. A hit means the model correctly predicted the ramp event within +/- 6 hours.

6. References

AWEA, 2009: 20% energy by 2030. Available at www.awea.org.

Bradford, K. T., R. L. Carpenter and B. Shaw, 2010: Forecasting Southern Plains wind ramp events using the WRF model at 3km. *Ninth Annual Student Conference*, Atlanta, GA, Amer. Meteor. Soc., [Available online at http://ams.confex.com/ams/90annual/techprogram/paper_166661.htm].

Carter, K. C., A. J. Deppe, and W. A. Gallus. Jr., 2011: Simulation of Nocturnal LLJs with a WRF PBL Scheme Ensemble and Comparison to Observations from the ARM Project. *24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., [Available online at http://ams.confex.com/ams/91Annual/webprogram/Paper179913.html].

Freedman, J., M. Markus, and R. Penc, 2008: Analysis of West Texas wind plant ramp-up and ramp-down events. *AMS Truewind Report*. 250-278.

Francis, N., 2008: Predicting Sudden Changes in Wind Power Generation. *North American Windpower*.