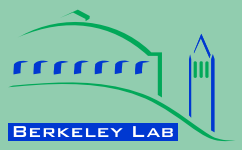
LBNL-XXXX



ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Analysis of Wind Power and Load Data at Multiple Time Scales

Katie Coughlin and Joseph H. Eto

Environmental Energy Technologies Division

December 2010

The work described in this report was funded by the Federal Energy Regulatory Commission, Office of Electric Reliability. The Lawrence Berkeley National Laboratory is operated by the University of California for the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Disclaimer

LBNL-XXXXX

Analysis of Wind Power and Load Data at Multiple Time Scales

Katie Coughlin and Joseph H. Eto

Ernest Orlando Lawrence Berkeley National Laboratory
1 Cyclotron Road, MS 90R4000
Berkeley CA 94720-8136

December 2010

The work described in this report was funded by the Federal Energy Regulatory Commission, Office of Electric Reliability. The Lawrence Berkeley National Laboratory is operated by the University of California for the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Abstract

In this study we develop and apply new methods of data analysis for high resolution wind power and system load time series, to improve our understanding of how to characterize highly variable wind power output and the correlations between wind power and load. These methods are applied to wind and load data from the ERCOT region, and wind power output from the PJM and NYISO areas. We use a wavelet transform to apply mathematically well-defined operations of smoothing and differencing to the time series data. This approach produces a set of time series of the changes in wind power and load (or "deltas"), over a range of times scales from a few seconds to approximately one hour. A number of statistical measures of these time series are calculated. We present sample distributions, and devise a method for fitting the empirical distribution shape in the tails. We also evaluate the degree of serial correlation, and linear correlation between wind and load. Our examination of the data shows clearly that the deltas do not follow a Gaussian shape; the distribution is exponential near the center and appears to follow a power law for larger fluctuations. Gaussian distributions are frequently used in modeling studies. These are likely to over-estimate the probability of small to moderate deviations. This in turn may lead to an over-estimation of the additional reserve requirement (hence the cost) for high penetration of wind. The Gaussian assumption provides no meaningful information about the real likelihood of large fluctuations. The possibility of a power law distribution is interesting because it suggests that the distribution shape for of wind power fluctuations may become independent of system size for large enough systems.

Acknowledgments

The work described in this report was funded by the Federal Energy Regulatory Commission, Office of Electric Reliability. The Lawrence Berkeley National Laboratory is operated by the University of California for the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

The Principal Investigator for the overall project is Joseph H. Eto, Lawrence Berkeley National Laboratory.

For their review and helpful comments on a draft of this report, the authors would also like to thank: Howard Illian (Energy Mark Inc.), Michael Milligan (National Renewable Energy Laboratory), Andrew Mills (Lawrence Berkeley National Laboratory), Phillip Price (Lawrence Berkeley National Laboratory), Michael Sohn (Lawrence Berkeley National Laboratory), and Yih-Huei Wan (National Renewable Energy Laboratory).

The study authors, alone, however, bear sole responsibility for technical adequacy of the analysis methods and the accuracy of the study results.

Table of Contents

List of Figures

List of Tables

1. Introduction

In this study we develop and apply new methods of analysis to high resolution time series data for wind power and system loads, with a focus on understanding the behavior of these signals at different time scales. As is well known, wind power output is highly variable and only partially controllable. Wind power can be forecast with some precision, and reduced if necessary, but not increased. In most areas system operators attempt to take whatever wind resources are available, so that from an operational perspective, wind is far more similar to load than to conventional generation resources. To date, wind resources have been integrated successfully into the electric grid at relatively low penetrations, and a key question is how to extrapolate wind power output and the correlations between wind power and load to the much higher penetration rates envisioned for the future.

System operations comprise different activities at different time scales; for example, frequency control at very short time scales of a few seconds to a minute, meeting regulation reserve requirements at the five to fifteen minute time scale, and hour-ahead and day-ahead refinements to the scheduled generation resources, primarily to compensate for forecast errors and outages. In systems where wind is used, reserve requirements must be based on the net load, defined as the instantaneous system load minus wind power. Important correlations between wind power and load exist because both are affected by weather patterns. A number of studies have examined whether the intermittent, "non-dispatchable" character of wind power may make it more difficult or more costly to maintain adequate system reserves. Integration studies look at hypothetical systems with much higher wind capacities than currently exist, and so are reliant on models to characterize the output of wind power plants. In a typical study framework (for example, GE Energy 2008) a meteorological model driven by historical weather data is used to estimate wind speeds and the resulting power output at wind plant locations, while load is represented by (possibly scaled) historical system load data for the same time period. Given the large number of approximations used in meteorological models, their output is only reliable at time scales of about one hour or longer (WWSIS 2008). To model system operations at shorter time scales, a statistical model for the variability of wind and correlation with load must be used. Most studies appear to assume that the wind-load correlations are adequately captured by using compatible historical data at the hourly time scale. Fluctuations in wind output at shorter time scales are often assumed to be Gaussian. In the present study, a close examination of the available data shows that these assumptions are not true; this may have important consequences for the accuracy of wind power integration studies.

The fluctuating character of wind power is often referred to as intermittency. It is important to note that intermittency is not the same as unpredictability. To the extent that intermittency in wind power is driven directly by weather, it is somewhat predictable, and could be managed in part through the use of near-term wind power forecasts (GE 2008, CAISO 2007). The presence of both deterministic and random components in the wind means that frequency distributions constructed from time series of wind power output should be strongly non-Gaussian, which is what we find in this study for all the regions examined.

As we are interested in examining the wind and load time series at different time scales, in this report we develop a wavelet-based approach (Burrus et al. 1998) to analyzing the data. Wavelets have been used to construct power spectra of short-term fluctuations in solar clearness index

(Woyte *et al*. 2007), but to our knowledge have not yet been applied to wind power data. Our application of wavelets is very simple, and no specialized background is needed to understand the analysis presented here. The wavelet transform is used to define a set of time series of the data at different time scales ranging from a few seconds to approximately one hour. Simple statistical measures (mean, standard deviation, linear correlation *etc*.) are then applied to these time series. We analyze wind power and load data from the ERCOT region, and wind power data from the PJM and NYISO areas. The data were collected and provided to us by staff of the Federal Energy Regulatory Commission (FERC) Office of Electric Reliability. The full FERC dataset included the California ISO, PJM, the New York ISO, Bonneville Power Authority, and ERCOT (Texas). Only ERCOT provided both load and wind data at high time resolution, so this data set is analyzed in detail. Results for wind only are given for NYISO and PJM. For the other areas, the time intervals at which the data were collected are too long for our approach to be useful.

The rest of this report is organized as follows: In section 2 we provide a detailed description of the algebraic steps used to process the data, define our statistical metrics, and present some illustrative time series examples. Section 3 provides detailed results for the ERCOT data (the results for PJM and NYISO are given in the Appendix). In section 4 we discuss the relationship between random processes, distribution shapes and the physical and engineering properties that need to be modeled in system integration studies. We present a simple method for analytically defining the shape of the distributions of wind power fluctuations, and provide evidence that suggests that these have an exponential shape near the center (*i.e*. for small magnitude fluctuations) and a power law in the tails. The behavior in the tails is particularly interesting, as a power-law distribution is typical of a so-called "scale-free" random process. For such a process, the distribution shape doesn't change under a re-scaling of the independent variable. In the current context, this means that the distribution shape would remain the same under an increase of the magnitude of total wind capacity; hence, it would be legitimate to extrapolate the results to much higher wind penetrations. Unfortunately, due to the limited amount of data currently available, fits to the distribution shape are very noisy and it is not possible to confirm these hypotheses. However as more data become available the validation is straightforward.

2. Analytical Approach

Wavelet transforms (Burrus et al. 1998) are designed to provide a mathematically rigorous technique for examining data at different scales of resolution. In the application presented here[1], algebraically simple operations are used to create smoothed version of the time series at a longer time scale, and a second time series which captures the "details" (also called differences or fluctuations) that have been removed by the smoothing. The operation is illustrated in Figure 2-1. Beginning with a series of values x(j) at the initial resolution T0, the smoothed signal y(k) is constructed by taking the simple average of adjacent values, while the "detail" signal a(k) is defined as the difference of adjacent values. The first pair of values [x(1), x(2)] define y(1) and a(1); the second pair [x(3), x(4)] define the values y(2) and a(2), *etc*., so the resolution of the two new time series y(k) and a(k) is T1 = 2*T0. Note that the series a(k) gives the change in the value of x over the original time interval T0, recorded at intervals of 2*T0.

The smoothing/differencing operation is repeated on the signal y(k), to create the two series z(m) and b(m), which have a time resolution of T2 = 2*T1 = 4*T0. The series z(m) is the smoothed signal while b(m) contains the details. The next iteration of this process would use z(m) to produce a smoothed series and a series of details at time scale T3 = 2*T2. In this way, with each iteration, the time interval between adjacent values in a series doubles, and the total number of elements in the time series decreases by a factor of two. For this analysis, the process is terminated when the time resolution reaches a scale of approximately one hour. Given the strong diurnal pattern in both wind and load data, and the underlying concern with daily system operations, continuing to longer time scales does not seem particularly useful. To capture seasonal effects and facilitate the algebraic steps, the data are segregated into four sets of 96-day seasons. As the initial time series data are quite high resolution (with T0 of four, five or six seconds), it takes nine iterations to go from the initial time resolution to the final hourly scale.

[1] This approach is equivalent to using the Haar wavelet decomposition.

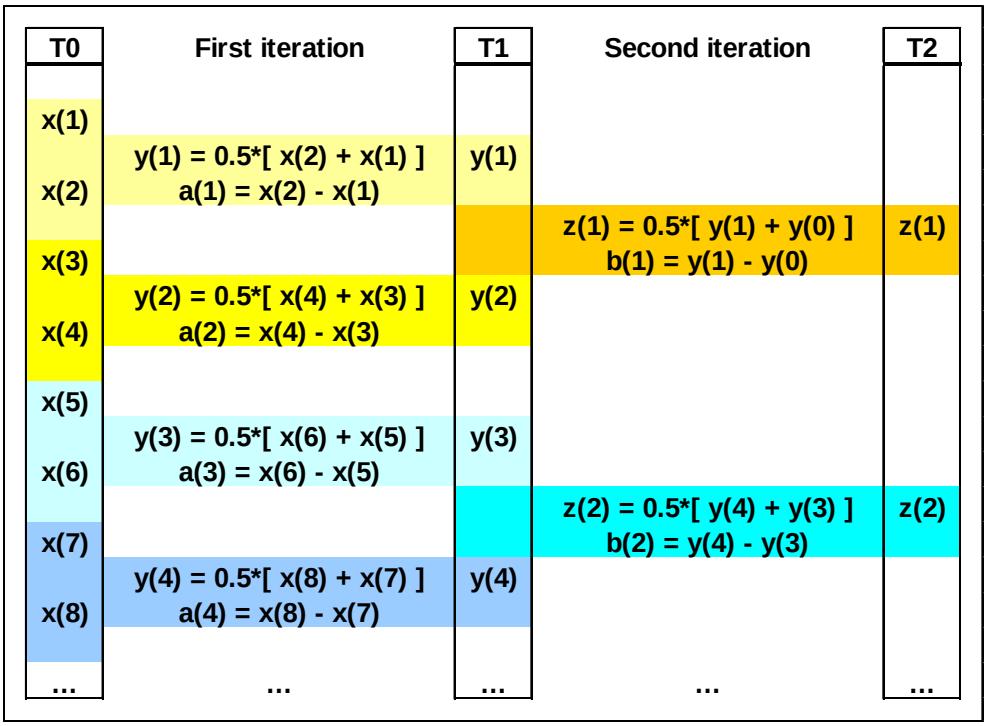


Figure 2-1. Illustration of the smoothing/differencing operations used.

2.1 Sample Data

Daily wind power profiles, and the results of the wavelet approach to smoothing and differencing, are illustrated in Figures 2-2, 2-3 and 2-4. Figure 2-2 shows the total ERCOT wind power output for the week beginning August 1, 2008 (this date was selected arbitrarily). The data begin at 12 midnight, and the horizontal axis is an index giving the number of days since the beginning of the plot. The plot illustrates a number of common features of wind output: a fairly regular diurnal pattern with stronger winds at night, a high degree of day-to-day variability, and the occurrence of large ramps (the ramp rate is defined as the change in output per unit time). Management of these large ramps is a key problem in the integration of wind generation.

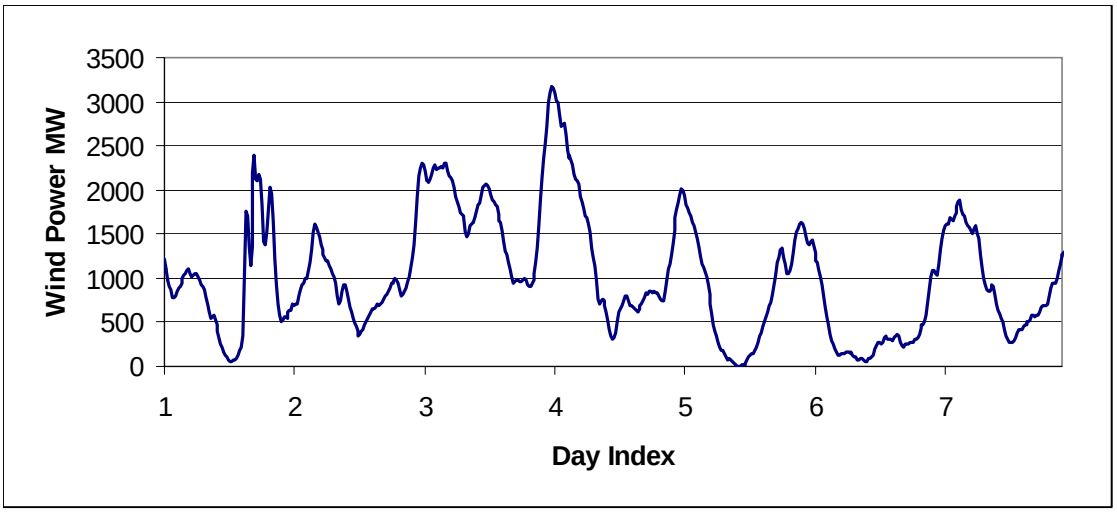


Figure 2-2. Total wind power output for one week beginning August 1st 2008

The next two figures illustrate the results of applying the smoothing and differencing described above. Figure 2-3 shows the smoothed signal, and Figure 2-4 the details. Each plot includes two different time-scales, 256 seconds (a little over 4 minutes, red triangles) and 1024 seconds (about 9 minutes, blue diamonds). The plots show the data for August 1st 2008. The horizontal axis is an hour index that indicates the number of hours since the beginning of the data series. In each figure, the upper plot shows the full 24 hours, while the lower plot focuses in on a few hours in the middle. In Figure 2-3 the two time-scales are barely distinguishable in the upper (24-hour) plot, but the lower (6-hour) plot shows both clearly. Figure 2-4 shows the differences that are the counterparts to the smoothed signals in Figure 2-3. Note that the magnitude of the differences is larger at the longer time-scale. This property is a consequence of serial correlations in the data; the differences at longer time scales can be thought of as resulting from the accumulation of differences at shorter time scales, as is illustrated clearly in the lower plot.

2.2 Statistical metrics

The time series of differences (the a(k) and b(m) above) define the "step changes" (Wan 2004) or "deltas" (GE Energy 2008) in the signal at different time scales. We will use the term "deltas" in the rest of this report. Because they define a change in magnitude over a given period of time, these deltas are equivalent to ramps, *i.e.* changes in wind or load that need to be dynamically managed to maintain system reliability. We will examine in detail the statistical properties of the deltas, including

- The simple average, denoted $\langle x \rangle$
- The root mean square (RMS), which is the square-root of $\langle x^2 \rangle$
- The standard deviation, defined as the square root of $\langle x^2 \rangle - \langle x \rangle^2$
- Frequency distributions (histograms)

All statistical metrics are presented as a function of the time scale, denoted $T_n = 2^n * T0$.

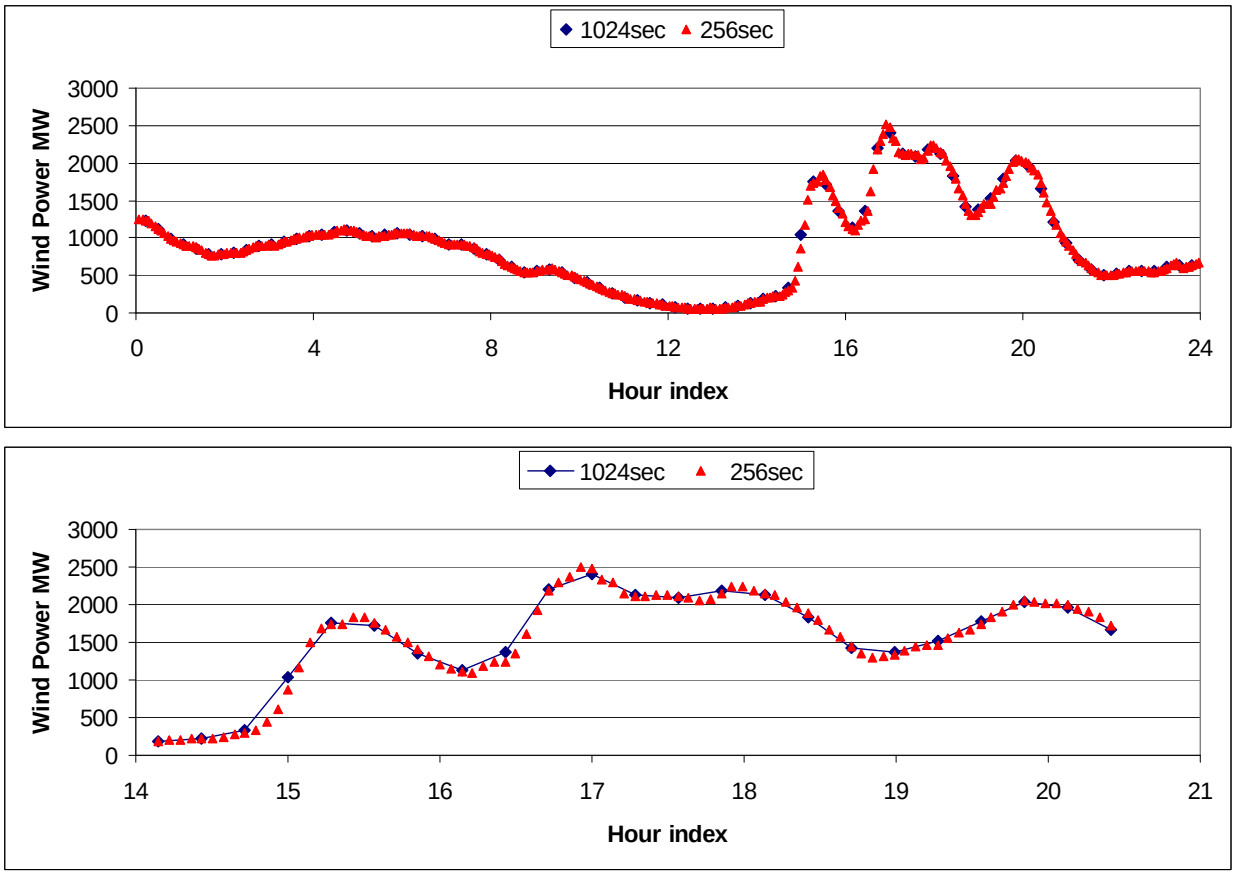


Figure 2-3. Total wind power (August 1st 2008). Each plot shows two time scales: 256 seconds (red) and 1024 seconds (blue). The top figure shows 24 hours and the bottom figure just over 6 hours.

We also look at correlations between the wind power deltas and the load deltas, and correlations between wind deltas at different time scales. The latter are important in trying to understand the degree to which a given ramp rate is likely to persist. Ultimately, any patterns that exist in either wind or load are driven by weather. Wind is directly affected by temperature gradients, so there is always some daily and seasonal pattern to the wind power output profiles. The strongest temperature gradients are aligned with the rising and setting of the sun, so generically one can expect a general pattern of wind pick-up and drop-off in either the morning or evening hours. As human behavior also follows the sun, loads too show a consistent pattern of ramping up and down in morning and evening periods. The degree to which wind ramps and load ramps move together or in opposite directions is an important question for system operators. If, for example, wind power drops over the morning period when load is ramping up, the addition of wind capacity to the system may require an increased ramp up of conventional resources. In this study, the correlations between wind and load in the ERCOT region are quantified using a simple linear correlation coefficient. This coefficient is calculated for all hours and for specific periods covering the morning ramp-up, afternoon peak period, and evening ramp-down.

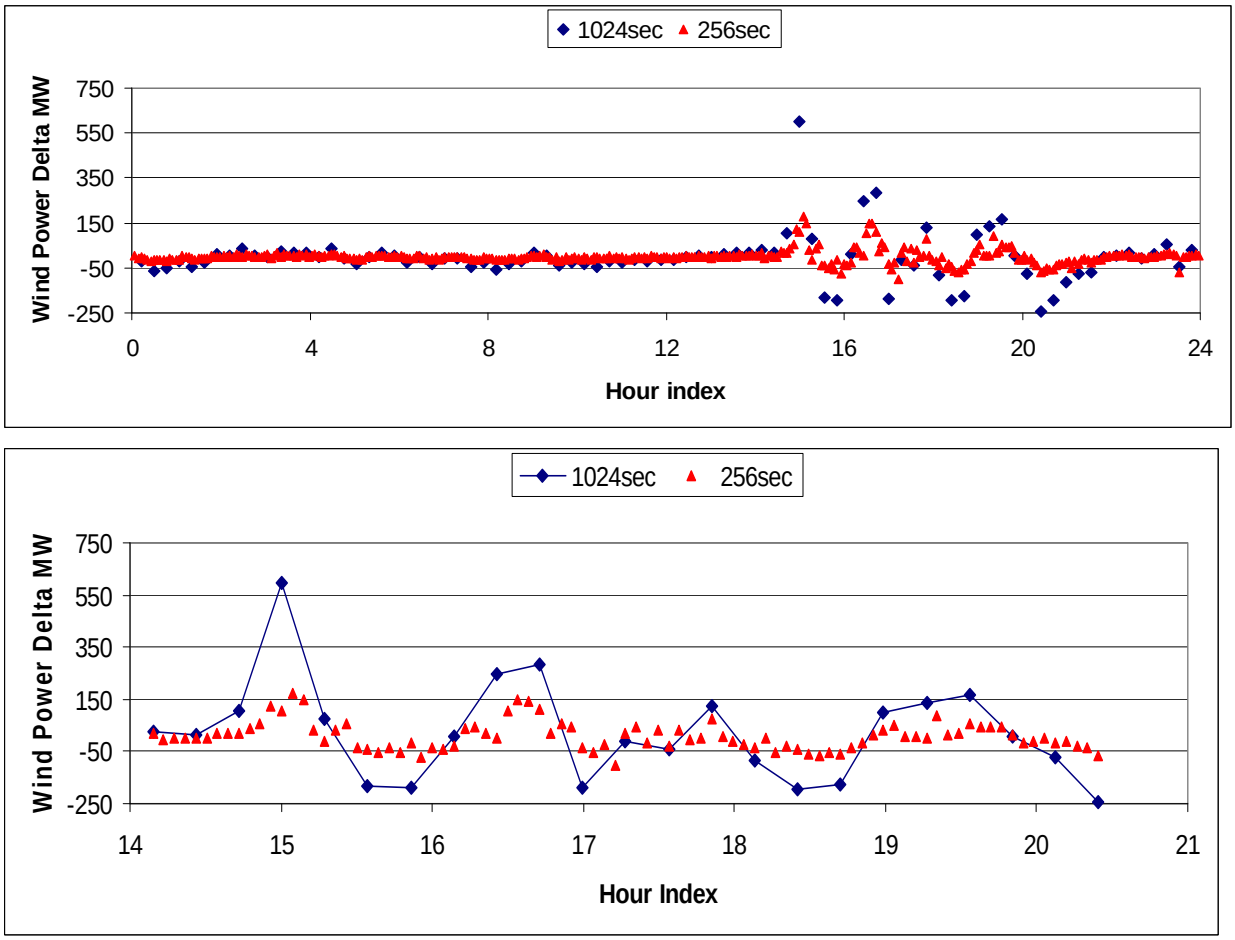


Figure 2-4. Wind power deltas (data from August 1st 2008), plotted as in Figure 2-3.

Serial correlation is also important to system operations. Serial correlation refers to the existence of a dependent relationship between the value of a signal at time *t* and its value at times earlier than *t*. For wind power, serial correlations occur due to the advection of large scale weather patterns across the wind plant site. Many decades of detailed study have shown that, while the atmosphere is strongly turbulent, it also exhibits persistent organized patterns at many spatial scales. Medium to large scale patterns lead to spatially correlated fluctuations across wind power plants located in the same area; this will appear in time series data as serial correlation. Operationally, serial correlation is important because it affects the duration of wind power ramps.

For system operations, both the magnitude and the duration of a ramp are important. If a ramp of a given size persists for a long time, the system operator may have to line up additional reserves to keep the system in balance, In a study of data for Minnesota, Ela and Kemper (2009) distinguish between ramping events which persist for on the order of an hour or more and lead to a substantial change in the MW provided by wind, and "false ramps" where the same ramp rate occurs but with a duration of 15 minutes or less. An important statistical question is whether the beginning of a true ramp event can be reliably distinguished from a "false ramp"; this will clearly require a precise analysis of subtle temporal patterns in the data. One way to approach this

problem is by looking at the linear correlation between wind power deltas of different time scales. During long-term ramping events, the deltas at all time scales will have the same sign and show a strong correlation. Conversely, for short-term ramps the correlation between long and short-time scale deltas will be weaker. Some sample calculations of this metric are presented for the ERCOT data. The persistence of ramps is also quantified by calculated the probability of occurrence of a sequence of *n* deltas of the same sign, as a function of both *n* and the time scale, and by looking at the total MW change associated with a series of length *n*.

The wavelet transform differs from the approach typically used in wind power studies in a way that is particularly significant for the analysis of serial correlations. Most studies use a simple difference that defines a delta at each time step as the difference d(j) = x(j)-x(j-1). With this definition, d(j-1) and d(j) both depend on x(j-1); this will introduce a degree of serial correlation into the d(j) series that might otherwise not have been there. Smoothed signals and deltas at longer time scales are typically defined using a running average, which also introduces an algebraic dependence between successive values. These algebraic dependencies can potentially lead to misrepresentation of the different correlations in the data. Conversely, the mathematical theory underlying the wavelet transform guarantees that the information in the series constructed for different time scales is algebraically independent; hence, any statistical correlations observed can be attributed to real correlations in the processes that generate the data.

3. Results for ERCOT

The ERCOT data cover one full year, from March 1 2008 through February 28 2009, at four-second time resolution. The wind data consist of a single time series, equal to the sum of the instantaneous output from all on-line wind resources. The American Wind Energy Association maintains a database of capacity additions, with totals updated by quarter (AWEA 2010). Examination of their data shows that approximately 2600 MW were added in Texas over the period April 1 2008 to March 31 2009. The capacity information is summarized in Table 3-1, which also shows the monthly average wind energy output calculated from the aggregate time series provided by ERCOT. The fact that total system capacity is expanding over the study period means that the aggregated time series data do not represent a statistically stationary system, which may affect some of the numerical results. This problem is not as severe as it may seem, because we divide the data into seasons and examine each season separately, so only the capacity additions during each 3-month sub-period need to be corrected for (the seasons are indicated in Table 3-1 by the color coding in the Month column). If the exact online dates for each project were available, we could scale the data for each month by total system capacity, which would reduce (but not eliminate) any numerical impact of system expansion on the statistics. However, total capacity additions are only available by quarter, and the quarters are not quite aligned with the seasons used here. Hence, the AWEA data could only be used to provide an approximate correction for capacity additions. In each quarter the capacity expansion is about 5-10%; the monthly values of the average wind output show that weather-induced variations in output are much larger. This suggests that the effect of system expansion on the numerical results should be relatively small, and as we're primarily interested in the qualitative behavior of the data for this study, we have not attempted to correct for capacity additions here. In general, system expansion effects can be corrected for if time series data disaggregated to the wind plant level are available.

Table 3-1. Average wind output and capacity in MW for the ERCOT region over the study period

Year	Quarter	Month	Avg Wind MW	MW Added	Total MW
2008	1	Mar	1973		5226
2008	2	Apr	2016		
2008	2	May	2092		
2008	2	Jun	2233	288	5514
2008	3	Jul	1512		
2008	3	Aug	776		
2008	3	Sep	845	693	6207
2008	4	Oct	1743		
2008	4	Nov	1908		
2008	4	Dec	2444	820	7027
2009	1	Jan	2148		
2009	1	Feb	2486		
2009	1	Mar		789	7817

For this data series, the maximum annual wind power output is about 4780 MW, which occurs in January 2009. Table 3-2 shows the average magnitude of wind power output, load, and the ratio of wind to load by season. Wind tends to be higher in the winter when loads are lower. The winter season is 96 days spanning December, January and February ("djf"), spring spans March,

April and May ("mam"), summer is June, July and August ("jja) and fall is September, October and November ("son").

Table 3-2. Average wind and load for ERCOT by season

Season	Months	Wind	Load	Net Load	Wind/Load
winter	dfj	2332	31266	28935	7.5%
spring	mam	2039	33307	31268	6.1%
summer	jja	1522	43706	42184	3.5%
fall	son	1509	32549	31041	4.6%

3.1 Statistics of the Deltas and Ramp Rates

Table 3-3 shows the RMS values of the wind, load and net load deltas at the full set of time scales $T_n = 2^n * T0$, n=1,2,…,10. The longest time scale is 4096 seconds or about one hour and eight minutes. The average value of the deltas is approximately zero, as fluctuations of opposite sign tend to cancel out. Hence, the RMS is approximately equal to the standard deviation for this signal, and can serve as an estimate of the fluctuation size irrespective of sign. At very short time scales of 8-32 seconds the deltas do not depend on T_n; around the 30-second time scale there is a change in behavior and the RMS values start to increase with T_n. This implies that for very short time scales the deltas behave like truly random noise, but for time scales on the order of a 30 sec to one minute and higher, serial correlations lead to an accumulation of fluctuations of the same sign. The table also shows what we call the "equivalent ramp rate" for a given delta. Let dW_n be the RMS value of the wind delta at time scale T_n. The equivalent ramp RW_n at this time scale is defined as the rate of change dW_n / T_n normalized to a one-hour time scale:

$$RW_n = dW_n * (T_n / \text{one hour})$$

Note these equivalent ramps do not correspond to real ramps; the ramp rate would have to last an entire hour to achieve an actual ramp of this size (the actual ramp at each time scale is just the delta divided by the time step). However, using equivalent ramp rates facilitates the comparison of magnitudes across time scales. The shorter the time scale, the larger the equivalent ramp. The trend with time scale is opposite to that of the deltas. Measured as equivalent ramps, the fluctuations become relatively insensitive to time scale for T_n larger than 256s (about 4 minutes).

Table 3-3. RMS values of the wind, load and net load deltas by season and time scale

T (s)	season	wind rms delta	load rms delta	net load rms delta	wind rms ramp	load rms ramp	net load rms ramp
8	winter	4.5	40.6	40.9	4086	36558	36801
16	winter	5.5	30.6	31.1	2480	13761	13995
32	winter	7.8	26.2	27.0	1755	5888	6082
64	winter	9.8	27.3	28.3	1103	3068	3183
128	winter	15.4	38.3	40.2	863	2153	2260
256	winter	25.4	65.4	69.2	714	1838	1946
512	winter	44.3	119.0	126.3	623	1674	1776
1024	winter	78.2	226.4	238.7	550	1592	1678
2048	winter	136.0	436.4	456.5	478	1534	1605
4096	winter	223.4	842.3	870.5	393	1481	1530

The data in Table 3-3 are for the winter months when wind power is at a maximum. All four seasons are shown in Table 3-2. The plots show both wind and load fluctuations, all seasons, as a function of time scale, measured as deltas on the left and equivalent ramps on the right. At each time scale, the series of points lined up on the vertical represents data for the four seasons. In the ERCOT region (Texas), wind power tends to be lowest in fall and highest in winter whereas loads are lowest in winter and highest in summer. Figure 3-1 also shows that the load seasonal variability is somewhat larger than the wind seasonal variability.

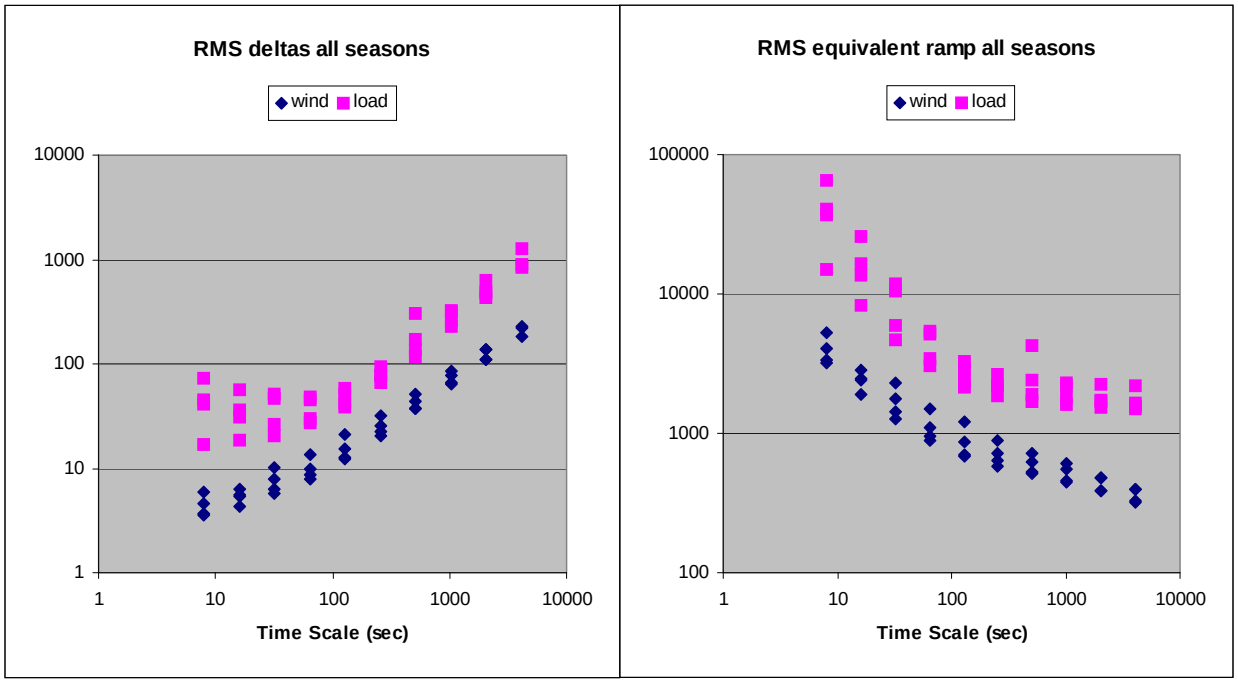


Figure 3-1. Load & Wind: RMS deltas and ramps by time scale

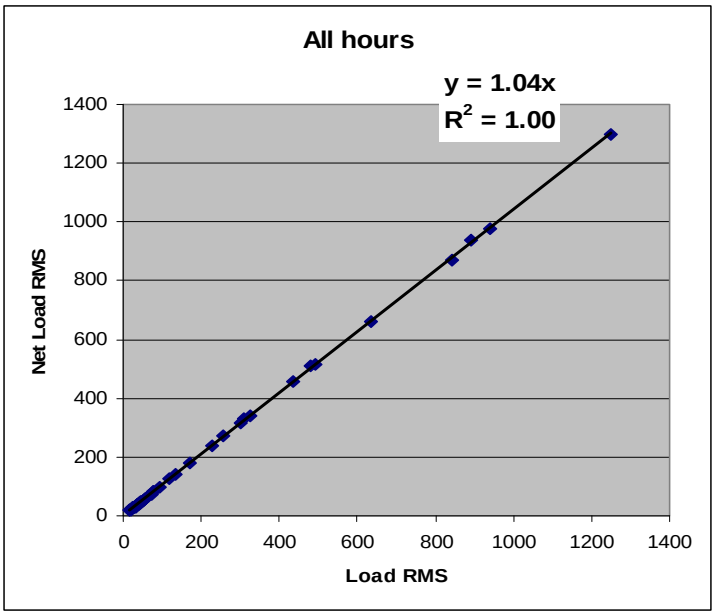


Figure 3-2. Variability of net load vs. load

Wind integration studies are generally interested in whether the variability of net load differs from that of load, where variability is defined as the RMS value of the deltas. When wind and load are anti-correlated (*i.e.* wind tends to be high when load is low and vice versa) the variability of net load will be higher than for load alone. In a study of the ERCOT region (GE Energy 2008), historical load data were combined with simulated wind power output for various levels of wind penetration, and the variability of net load was found to be consistently higher than that of load. The increased variability of net load over load was also found to depend on the time scale used to define the deltas, with higher variability at longer time scales. The results of a similar analysis using our data set are shown in Figure 3-2. In this figure, the RMS values for net load are plotted against the RMS values for load, for all seasons and time scales; each point corresponds to one particular season and time scale. Unlike the ERCOT study, the data show no dependence on time scale; instead, the increased variability of net load is approximately constant for all seasons and time scales.

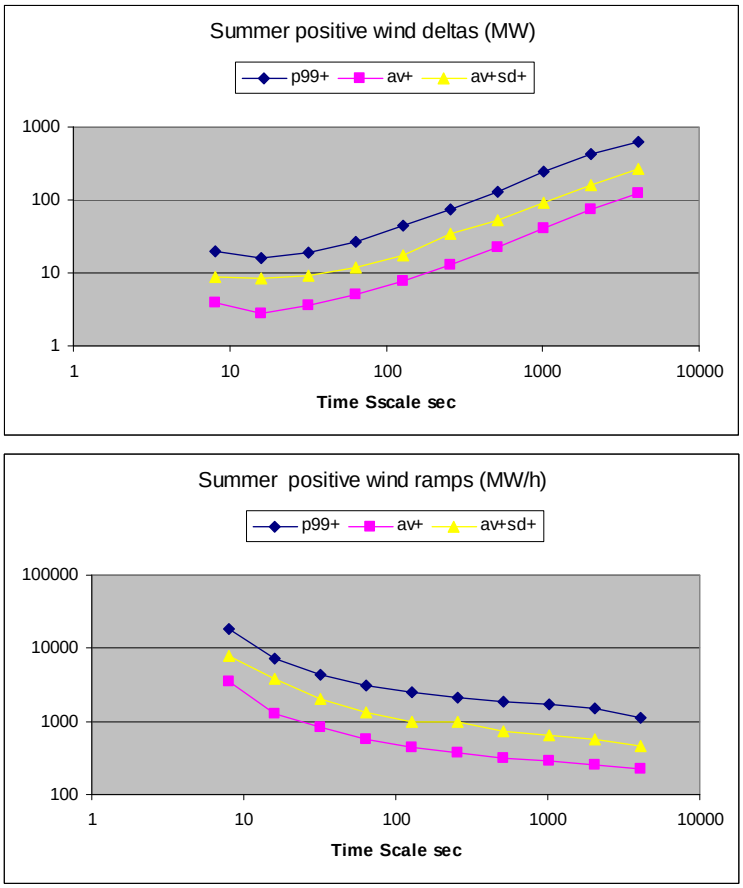


Figure 3-3. Wind: Positive deltas and ramps by time scale

To look at the statistics of positive and negative deltas separately, the data were segregated according to the sign of the delta. For wind data, the positive and negative deltas show very similar behavior. In Figure 3-3 the average (purple), average plus one standard deviation (yellow), and 99th percentile values (dark blue) for wind deltas with positive sign are shown, both the absolute values and represented as ramp rates. The data season is summer (JJA). The 99th percentile values are approximately two standard deviations away from the mean values. In Figure 3-4, we plot the fraction of all wind power deltas that are within one standard deviation (sigma) of the mean. For a Gaussian distribution 68% of points values are within one sigma, whereas for the wind power data, the percent of values that lie within one sigma varies from 75% to 90%. The fraction decreases as time scale increases, indicating that the distributions are broader at longer time scales. Further analysis shows that 95% of all values are within two sigma and nearly 100% are within 3 sigma. These results are consistent with the analysis of Wan (2006) of data for Kentucky.

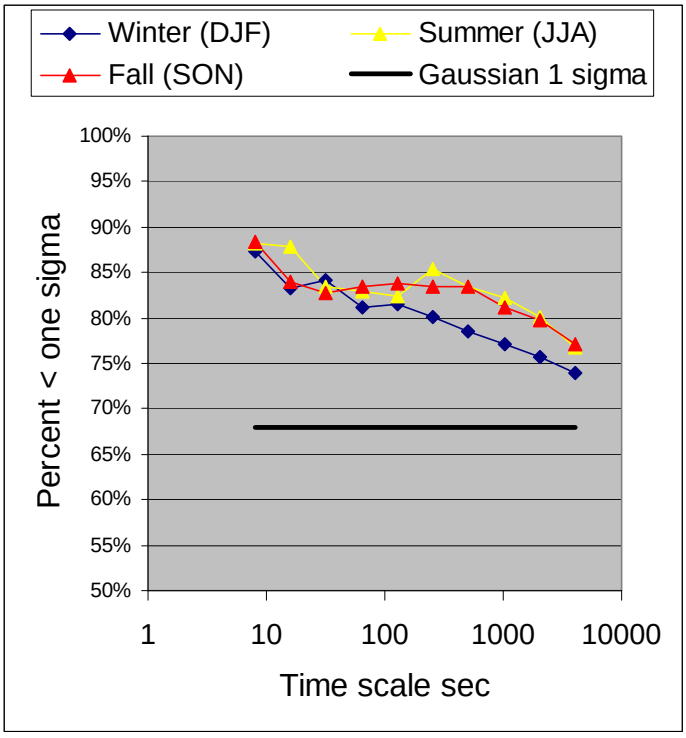


Figure 3-4. Percent of all deltas lying within one sigma of the mean for wind.

3.2 Distribution Shapes for Wind Power and Load Deltas

Frequency distributions for the wind power and system load deltas are presented in this section. These distributions are approximated by constructing histograms, where the data are sorted into bins of size k*B, where k is the bin index and B is the bin size. The bin size is chosen somewhat arbitrarily, to produce easily legible plots. A systematic approach to estimating distribution shapes is presented in section 4; here the plots are intended to be illustrative. Figure 3-5 and 3-6 show the wind power (red bars) and load (blue bars) data on the same plot, for summer and winter respectively, for a time scale of 128 seconds (about two minutes). The vertical axis in all these plots is logarithmic. The wind power delta distributions are very concentrated at the center and approximately symmetric. The load delta distributions become much tend to be broader, less symmetric, and show a strong seasonal variation.

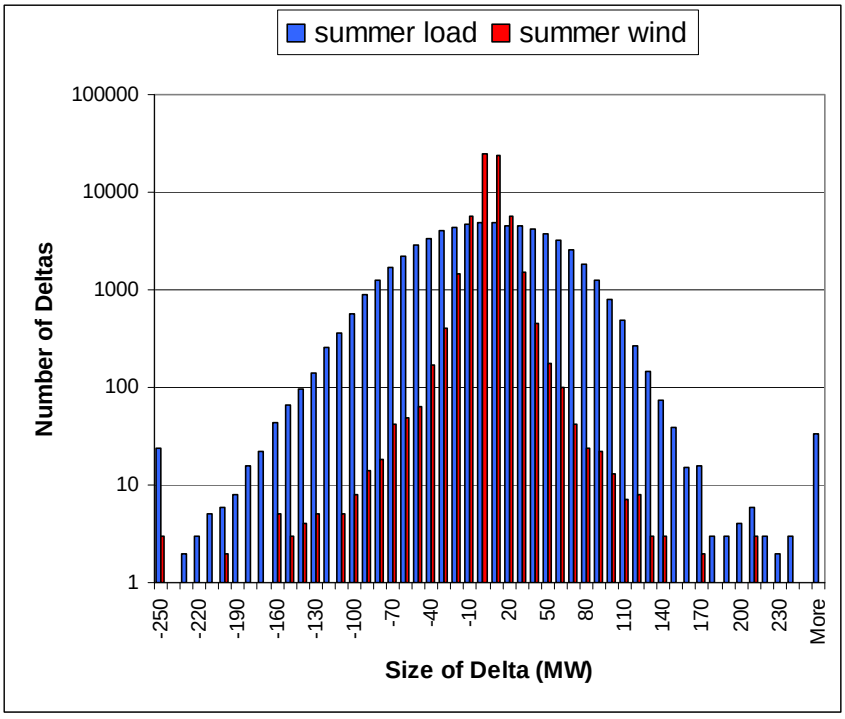


Figure 3-5. Frequency distribution of wind and load deltas for ERCOT, summer data.

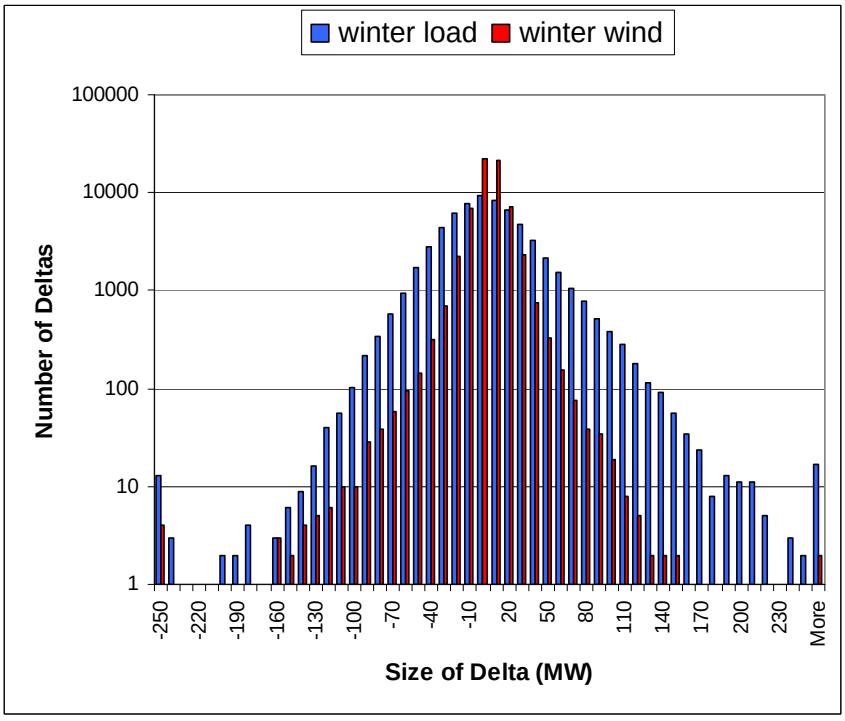


Figure 3-6. Frequency distribution of wind and load deltas for ERCOT, winter data.

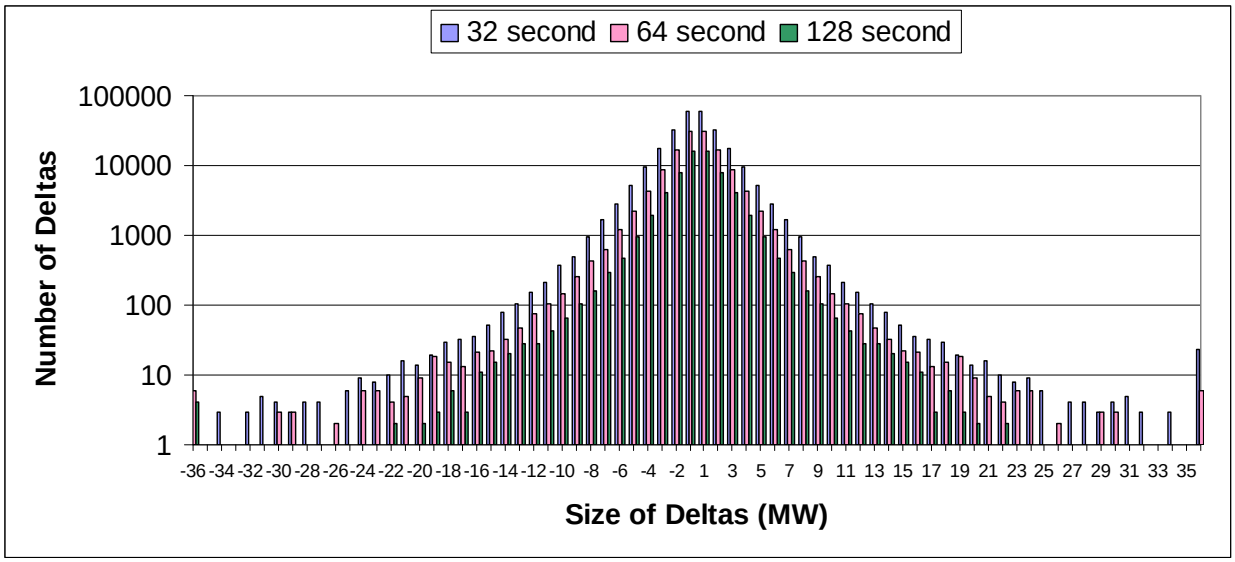


Figure 3-7. Frequency distribution of wind deltas at short time scales (winter).

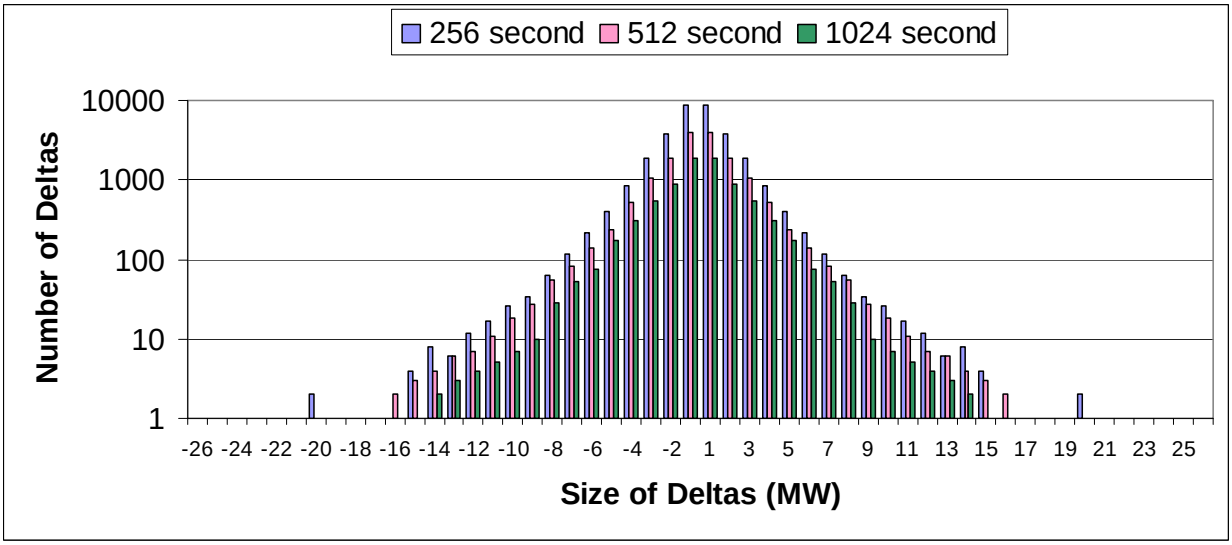


Figure 3-8. Frequency distribution of wind deltas at long time-scales

Figures 3-7 and 3-8 provide a view of how the distribution shapes for wind power deltas vary with time scale (these figures are presented in the Appendix for the PJM and NYISO data). In these plots, the bin index, rather than the bin boundary value, is plotted along the horizontal axis, so that distributions at different time scales can be easily compared. Figure 3-7 shows short time-scale data. The bin size is defined as four times the 99th percentile value divided by the number of bins (approximately 5MW). Figure 3-8 shows the long time-scales. In this figure the bin size is defined as three times the 99th percentile value divided by the number of bins (approximately 28MW). Long tails are particularly visible in the short-time scale distributions; they may also be present at long time scales but the smaller number of data points in the latter series means the distribution may not be filled out sufficiently to show them.

Overall, the distribution shape does not vary strongly with time scale. With a logarithmic vertical axis, a symmetric exponential distribution would have an outline consisting of two straight lines. This is very like what is seen in the plots for $|k| < \sim 7$. At larger values of k, the distributions seem to flatten and show fatter tails than an exponential. A closer examination of the distribution shape is provided in section 4.

3.3 Segregation of the data by time of day

Some periods may be more problematic from an operational perspective because of rapid increases or decreases in load; these include

- Morning ramp up: 5am-7am weekdays
- Evening ramp down: 10pm-12am weekdays
- Peak demand hours: 4pm-6pm weekdays

In this section we present various measures for the deltas restricted to these three periods. Table 3-4 shows the average load and average wind power for these three periods and for all hours, as well as wind as a percent of load.

Table 3-4. Wind, load and net load values segregated by time of day

64sec	5-7am				4-6pm			
Season	Wind	Load	Net Load	Wind %	Wind	Load	Net Load	Wind %
winter	2571	32748	30177	**7.9%**	1939	32473	30535	**6.0%**
spring	2322	30241	27919	**7.7%**	1849	39602	37752	**4.7%**
summer	1654	34860	33205	**4.7%**	1188	55203	54015	**2.2%**
fall	1794	28963	27169	**6.2%**	936	38967	38031	**2.4%**
	10pm-12mid				All Hours			
Season	Wind	Load	Net Load	Wind %	Wind	Load	Net Load	Wind %
winter	2730	30451	27722	**9.0%**	2332	31266	28935	**7.5%**
spring	2434	33259	30824	**7.3%**	2039	33307	31268	**6.1%**
summer	1890	42852	40961	**4.4%**	1522	43706	42184	**3.5%**
fall	1666	31167	29501	**5.3%**	1509	32549	31041	**4.6%**

Unlike the full dataset, for these short periods of time the average value of the load and wind deltas is not zero; the seasonal averages are presented in Table 3-5 for a few representative time scales. Ramp rates are presented as this is the quantity that is most relevant to operations. The early morning and late evening load ramps are always positive and negative respectively, independent of season and of time scales. The wind ramps in this data are negative in the early morning and positive in the late evening. As the wind and load ramps have opposite sign, this simple metric would suggest that the ramps in net load would be larger than load alone during these periods. For both wind and load, during the afternoon peak hours the sign of the ramp depends both on season and on time scale. This is partially due to seasonal changes in the time of sunset, which affects both loads and the temperature gradients that drive wind.

Table 3-5. Average values of the wind and load deltas for specific time periods

1-hr Ramp		5am-7am		4pm-6pm		10pm-12am	
Tsec	Season	load	wind	load	wind	load	wind
64	winter	3506	-36	1177	-73	-2132	110
64	spring	2851	-46	-28	43	-2882	72
64	summer	1601	-112	-644	91	-3733	141
64	fall	2881	-74	-39	11	-2818	69
512	winter	3495	-32	1178	-64	-2091	111
512	spring	2859	-51	-40	44	-2913	57
512	summer	1616	-116	-568	99	-3738	122
512	fall	2902	-71	-23	17	-2848	83
4096	winter	3398	0	1156	-55	-2109	103
4096	spring	2770	-11	-125	55	-2945	73
4096	summer	1605	-124	-567	117	-3751	128
4096	fall	2826	-49	-28	24	-2797	56

The intermittency in wind may be expected to increase the variability in load relative to net load. This is true to some degree as illustrated in Figure 3-9. This plot shows, for each season and time scale, the RMS value of the net load deltas vs. the RMS value for the load deltas during the hours 4-6pm weekdays. The trend line indicates that net load is 2% more variable than load in this metric; similar plots for the other periods show an even tighter straight line relationship. In the late evening net load is 4% more variable than load, while in the early morning net load is 2% more variable than load. As noted above for Figure 3-2 the increase in variability of net load over load does not show any dependence on time scale, or on season.

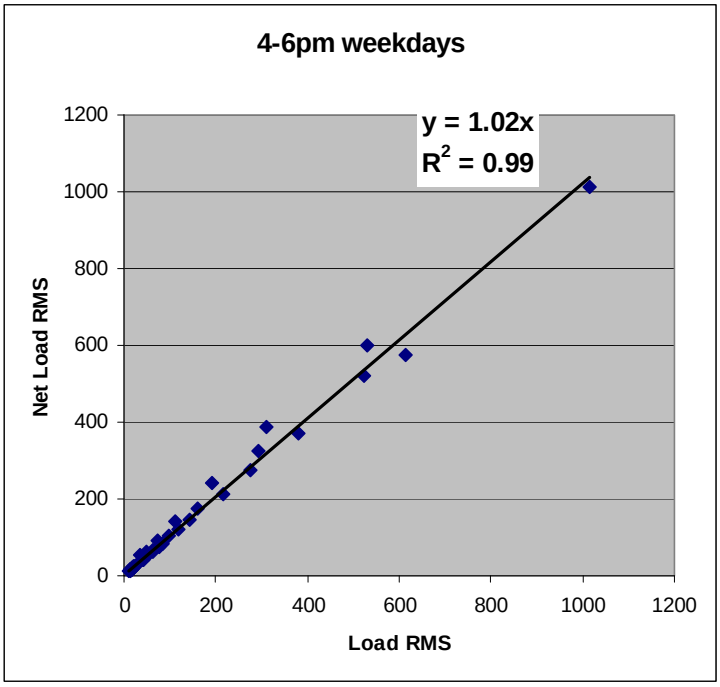


Figure 3-9. Variability of net load vs. load for the period 4-6pm weekdays (cf. figure 3-2)

To further investigate the correlation between wind and load during these periods, we calculate the uncentered linear correlation coefficient. For a given time scale, we define

- dW as the wind delta
- dL as the load delta
- rmsW as the RMS value of the wind deltas
- rmsL as the RMS value of the load deltas

The uncentered linear correlation coefficient is then

$$L = < (dW * dL) > / (rmsW * rmsL).$$

For this calculation, outliers have been removed from the data. These are defined as values which are greater than twice the 99th percentile value. Figure 3-10 shows the correlation coefficients for each of the winter and summer (the plots for spring and fall are qualitatively similar although they differ in detail). For comparison, the plots also show the correlation between wind and load for all hours together.

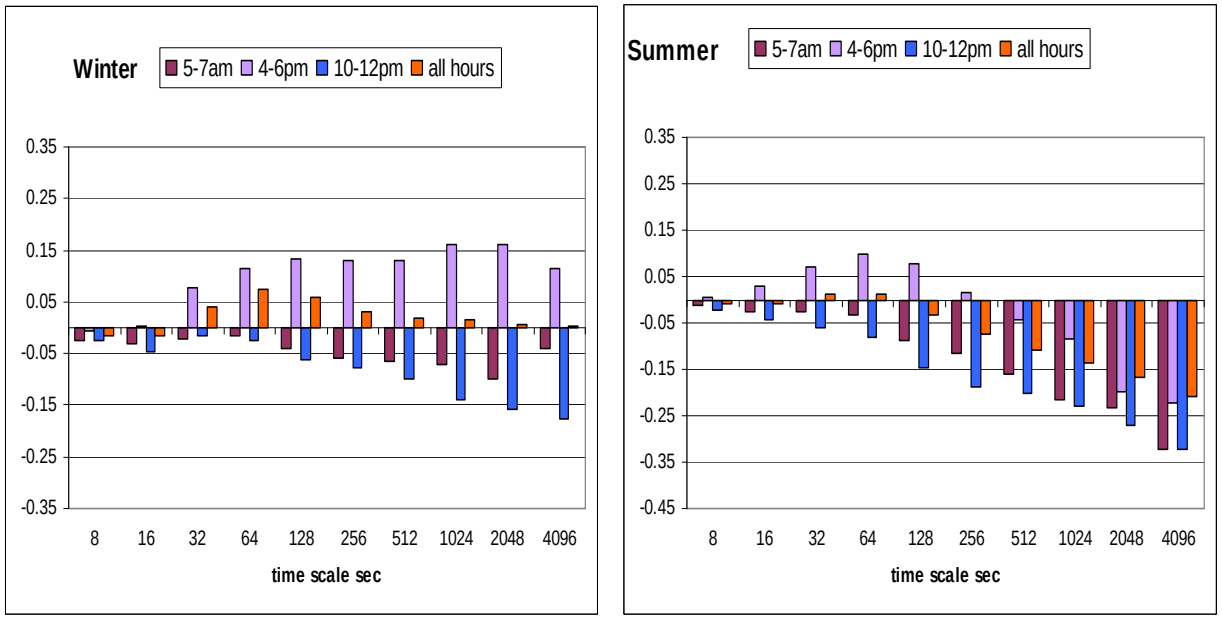


Figure 3-10. Linear correlation between wind and load deltas for summer and winter

The correlation coefficients are the same irrespective of whether deltas or ramp rates are used. These plots confirm that the wind and load ramps are negatively correlated in the early morning and late evening. During the afternoon peak, in winter they are positively correlated at time scales greater than half a minute; while in summer they become positively correlated at time scales on the order of a few minutes (256s). For very short time scales (< 30s) the correlations are weak for all periods and seasons.

It is not immediately obvious why the correlation coefficient would change sign as the time scale increases. As we have already noted, at longer time scales the deltas tend to accumulate into sustained intervals of changes with a given sign. It is therefore reasonable that correlations tend to increase with time scale, and are negligible at very short time scales where the fluctuations are truly random. In winter, higher wind (positive deltas) would tend to increase heat loss from buildings and thus increase space heating loads (and vice versa); in summer higher wind would tend to reduce space cooling loads (and vice versa). It therefore seems plausible that wind and load would be positively correlated in winter and negatively correlated in summer because they are both influenced by common large-scale weather conditions.

The patterns exhibited in these correlation features will vary by geographic region. The qualitative behavior of the correlation coefficients – the sign and order of magnitude by season and time scale, could provide a useful check on whether modeling studies are capturing the correlations between load and wind accurately.

3.4 Serial correlation

The accumulation of deltas of a given sign means that there are serial correlations in the data. These in turn could lead to sustained ramps in wind power output (of either sign) that can create difficulties for system operators. To quantify the degree of serial correlation, we use a simple metric that measures the probability that any given observation is part of a series of deltas with the same sign, as a function of the length of that series. For each element of the time series (indexed by i), we look at sequences of consecutive values x(i), x(i+1), … x(i+n) and count the number of times all *n* values are positive or negative. This count, divided by the total number of sequences of length *n* in the data, is an estimate of the probability of *n* consecutive positive or negative values. A series of length *n* implies the existence of series of all lengths less than *n*, so the probability of a series of a given length decreases with n. The results for wind data are shown in Figure 3-11 and Table 3-6. Figure 3-11 shows that sequences of positive and negative deltas have roughly the same dependence on time scale irrespective of sign. The probability of a sequence of length *n* is low at very long and very short time scales, and peaks near the center for time scales on the order of a few minutes. The fact that the probabilities decrease for longer time scales is significant – it implies an upper bound on the total MW change associated an extended ramp, *i.e*. a series of deltas with the same sign.

These sequences can be converted to ramp rates by summing the deltas to get the total change across the sequence. This calculation shows that the ramp rates do not depend strongly on time scale for time scales greater than one minute, and are insensitive to the value of *n*. Sample data are presented in Table 3-6 for wind power in winter. The ramp rate data converge to a slowly varying function of time scale that is approximately independent of series length. For time scales greater than a minute, each doubling of the time scale reduces the ramp rate by 10-15%. The total change in power output is the ramp rate times the duration of the ramp, so longer sequences do result in large total MW changes. Again, the fact that the probability of a series of deltas of a given sign peaks for a series length of 4 or 5, and that the equivalent ramp rates converge to a constant independent of series length, implies an upper bound on the total MW change associated with an extended ramp, at least on average. Further manipulations of the data along these lines could be used, for example, to calculate the probability of the extension of a ramp of length *n* to length *n+1*, but would require longer time series to produce useful results.

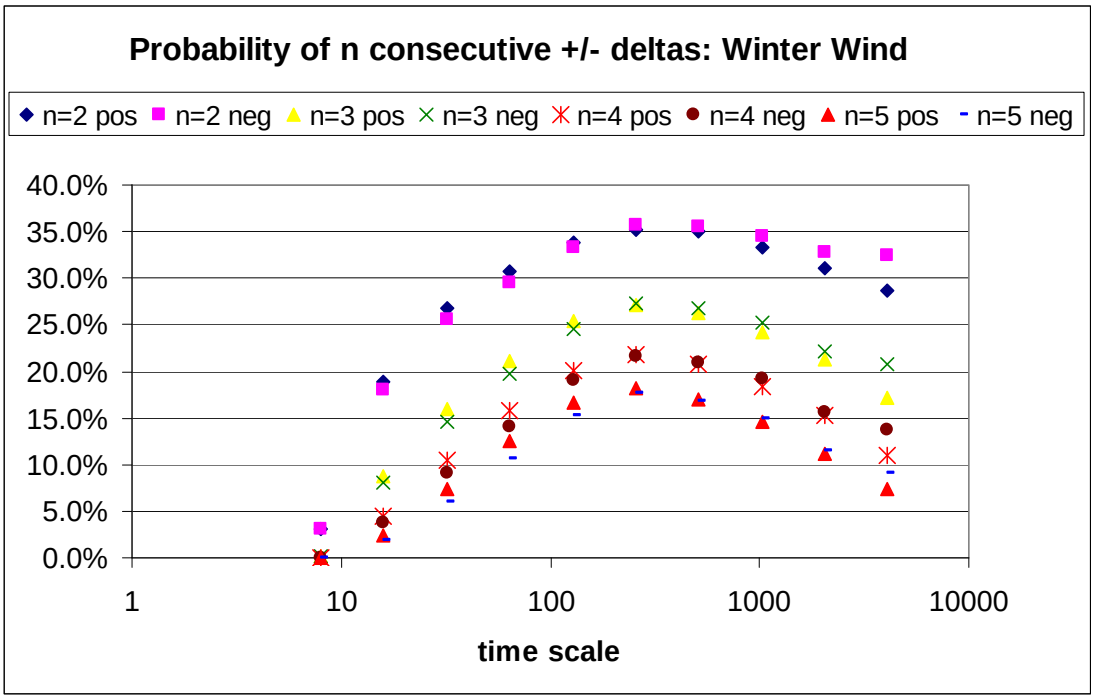


Figure 3-11. Probability of a sequence of deltas with the same sign as a function of time scale and sequence length

Table 3-6. Ramp rates associated with sequences of length n for winter wind

	positive deltas				negative deltas			
Tsec	n=2	n=3	n=4	n=5	n=2	n=3	n=4	n=5
16	1505	1503	1527	1562	-1562	-1569	-1603	-1659
32	998	1033	1072	1104	-1041	-1089	-1148	-1213
64	736	783	822	851	-762	-819	-867	-904
128	590	633	662	679	-595	-635	-660	-674
256	516	547	567	583	-501	-519	-530	-538
512	459	487	505	515	-438	-448	-454	-456
1024	417	439	452	455	-385	-393	-396	-399
2048	366	377	377	373	-330	-338	-339	-333
4096	317	325	336	324	-271	-271	-276	-281

As discussed in Section 2, correlation between the deltas at different time scales can be used to distinguish between short- *vs*. long-duration ramps. The correlation between the 256 second and 1024 second data are shown in Figure 3-12 as a function of the hour of the day. For comparison, applying this method to uncorrelated random data series leads to coefficients on the order of 0.01 to 0.1 The plots show a higher degree of correlation occurs in the morning and evening hours, and less in the afternoon hours. This suggests that extended ramps would be more likely in the morning or evening. The plots are somewhat noisy, as dividing up one year's worth of data by hour of day leads to rather sparse data sets. Longer time series are needed to determine whether this type of plot can provide useful operational information.

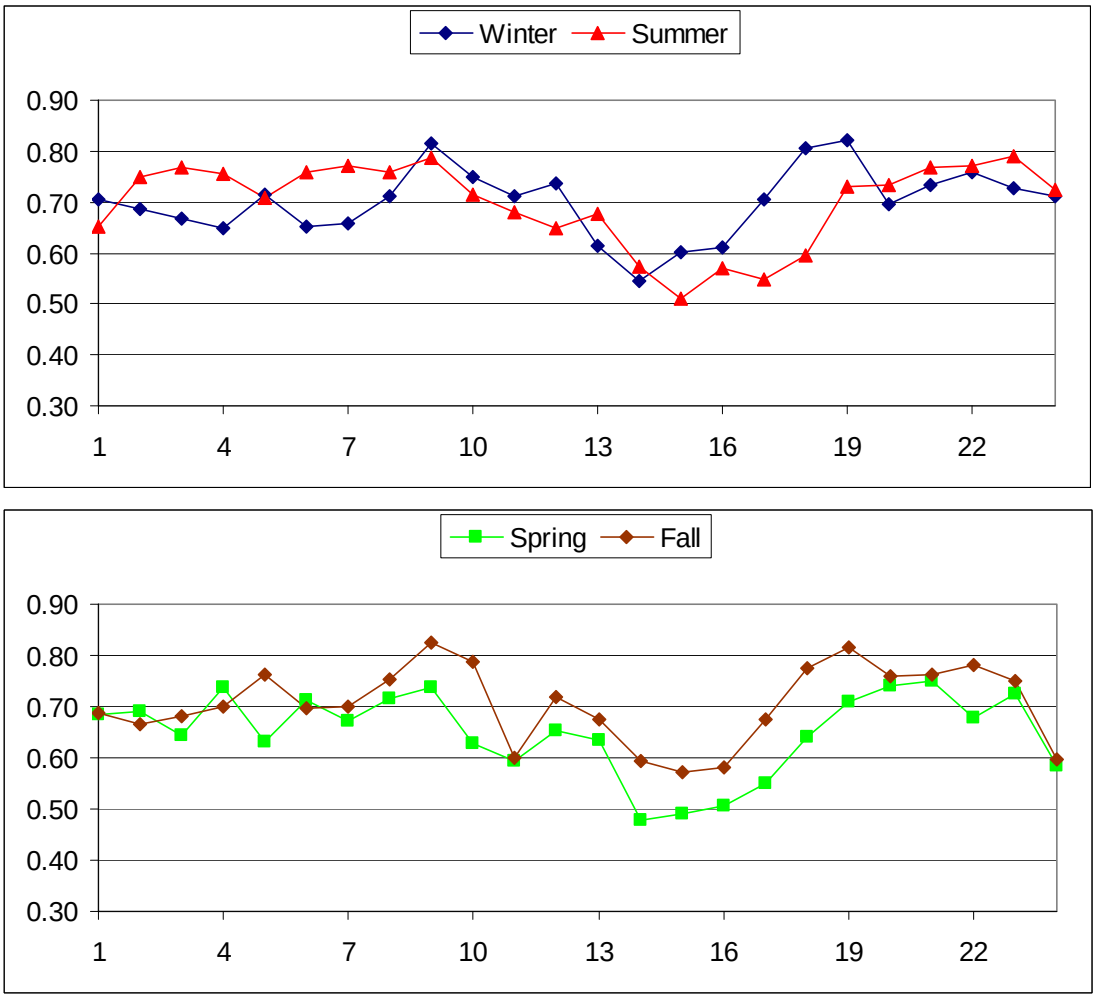


Figure 3-12. Correlation between wind deltas at timescales 1024s and 256s.

4. Discussion

4.1 Random Processes and Distribution Shapes

There are a few studies that have looked at the statistics of wind power deltas at different time scales (GE Energy 2008, Wan 2004, Wan & Liao 2006); these also seem to show distributions that are strongly peaked near the center, although it is not always clear from the way the data is presented. None of these studies has attempted to derive the empirical distribution shape from the data. In the next section we present an approach and apply it to the ERCOT data. Before getting into the details, it seems useful to review a few of the different physical processes that generate random signals having characteristic distribution shapes. There are a wide variety of random processes in both natural and engineered system, each of which generates a characteristic distribution function. In understanding complex systems, thinking about the processes that produce the random component of a time series can be very useful. We will argue that, based on the different temporal and spatial scales of processes that affect the wind power data, the distribution shape can have a different functional form in the center than it does in the tails.

Very loosely, a Gaussian distribution is generated by a random-walk process, in which successive steps are uncorrelated. The canonical example of a random walk is a sequence of coin flips. Since long series of random steps in the same direction are very rare, for Gaussian processes large excursions from the mean are very improbable and the distribution does not have long tails. Plotted against a logarithmic vertical axis as in the figures above, a Gaussian would look like an inverted parabola with a pronounced rounding at the center, qualitatively similar to the load delta distribution if Figure 3-5. This feature is not present in the distributions of wind power delta. Gaussian distributions may also result from the addition or averaging of a large number of *independent* random variables-- this is known as the Central Limit Theorem. However, the Central Limit Theorem does not apply here because the output of a collection of wind power plants is not equivalent to a set of independent random variables. On the contrary, their output can be highly correlated due to large-scale weather systems.

The log of frequency for the wind data more closely follows a straight-line profile as a function of the bin index k, which corresponds to an exponential distribution. Exponential distributions arise, for example, in queuing theory; the length of time between successive incoming calls to a central telephone exchange roughly follows an exponential shape. This distribution type allows for some degree of serial correlation, and also has the property of being “memory-less”. Loosely, this means that the statistical behavior of a series of a given length does not depend on when the series begins.

Both Gaussian and exponential distributions have a well-defined center (at zero for a symmetric shape) and a width that sets a scale for the distribution and reflects some aspect of the random process. In contrast, a power-law distribution has the property of being “scale-free”, *i.e*. it looks the same under a rescaling of the independent random variable. In physics, power law distributions often describe the fluctuations in “noisy” processes. The fluctuations in wind speed produced by a turbulent atmosphere are well known to obey a power law distribution over a broad range of scales, and the power spectrum of wind power time series data have been shown to have a similar shape (Apt 2007). If the distribution of observed wind power deltas from a collection of wind plants is truly scale-free, then the addition of more plants to the system will

not affect the distribution shape. The enlarged system can be modeled just by rescaling the magnitudes of the fluctuations.

Wind power output is influenced by different physical factors at different temporal and spatial scales, so it is reasonable to expect the functional form of the distribution shape to change as a function of the magnitude of the deltas.[2] In particular, relatively small changes in wind power output are due to spatially localized wind speed fluctuations that affect only one plant, while larger fluctuations must be due to broader, spatially coherent atmospheric patterns that simultaneously affect larger numbers of plants. The spatial correlation of weather patterns is driven partly by deterministic features that are more likely to produce large fluctuations than purely random processes. These features include local geography (mountains, lakes etc.), so there will be some spatial system below which the behavior of a collection of wind plants will not be generic. However, as the area covered by a set of integrated wind farms increases, the behavior of large wind deltas will eventually be determined by the typical spatial scale of atmospheric phenomena. This suggests that the distribution shape for wind power deltas will depend on the total system size[3] up to a point; once the system is large enough to be insensitive to the details of local geography, the distribution shape should become independent of system size. It would be very useful to wind integration studies if this hypothesis could be tested, and the distribution parameters estimated from data.

4.2 Empirical Distributions and Extreme Events

Empirical distributions are constructed from the time series data by binning values according to their magnitude. In this section we present a simple method for modeling the distribution shape from the data. The plots show that the distribution shape for wind power deltas are symmetric and centered at zero, so we assume these features in defining the empirical fits.

We define the constant bin width as w, the histogram index as k, and N(k) as the number of observations in bin k. The shape of the distribution can be investigated by looking at the behavior of N(k) as a function of k. To simplify the comparison with standard distribution shapes (such as Gaussian, exponential or power law), we calculate the function

$$r(k) = \log[\, N(k+1)/N(k) \,]$$

For unimodal distributions, the number of observations in each bin, N(k), decreases uniformly with the bin index. The function r(k) is a measure of the rate of decrease of the probability of finding an observation in bin k as k increases. For a Gaussian distribution, the log of the number of points in bin k is proportional to k^2, for an exponential distribution the log of N(k) is proportional to k, and for a power law distribution the log of N(k) is proportional to log(k). Some straightforward algebra leads to the relationships:

[2] Mathematically, a power law frequency distribution is not defined as the magnitude of the fluctuation tends to zero, so any process which follows a power law in the tails must follow some other distribution shape at the center.

[3] Here system size refers both to the spatial distribution of wind farms and the total MW.

- $r(k) = -(w/\sigma)^2 (k+1)$ for a Gaussian distribution with standard deviation σ
- $r(k) = -\alpha$ for an exponential distribution with rate parameter α
- $r(k) = -\beta/k$ for a power law distribution with exponent β

Qualitatively, for a Gaussian distribution the rate of decrease increases with k, for an exponential it is constant for all k, and for a power law the rate of decrease of the bin counts decreases with k. From the plots shown in section 3, the data appear to be consistent with a power law for larger k, and an exponential distribution near the center. The data do not appear to be consistent with a Gaussian distribution at any value of k. (This is also true for the PJM and NYISO data shown in the appendices). Figure 4-1 shows the function r(k), calculated for both wind power and load deltas at a time scale of 128 seconds, as well as illustrative curves for a Gaussian (black line) and power law (red line). As noted above, an exponential would appear as a straight line on this plot. This view of the data shows more clearly that the functional form of the distribution is different near the center than it is in the tails. For the load data, the shape looks Gaussian for k less than 6 or 7; for larger k the value of r(k) tends to a constant, indicating that the load delta distribution is exponential. For the wind data, the value of N(k) drops rapidly for the first few k; beyond that the data conform roughly to a power law but the fit is very noisy due to the relatively small number of observations in the outlying bins.

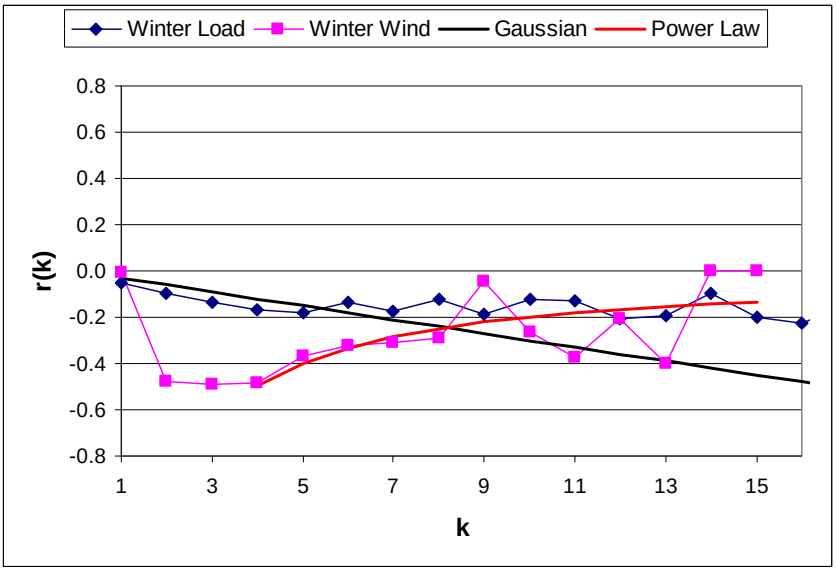


Figure 4-1. Plot of the function r(k) for 2-minute winter wind and load data

To examine the ERCOT data more closely, we assume the distribution is symmetric and sum the counts for bins labeled +/- k, which increases the number of observations in each bin and produces a somewhat smoother curve. The number of observations in each bin is plotted against the bin index on a log-log plot in Figure 4-2. The upper plot shows the wind power deltas for a long time scale (1024 seconds) and the lower plot the data at a shorter time scale (64 seconds). In each plot the data for winter and for all seasons together are shown. In this view, a power law should look like a straight line, which is roughly true for k larger than about six.

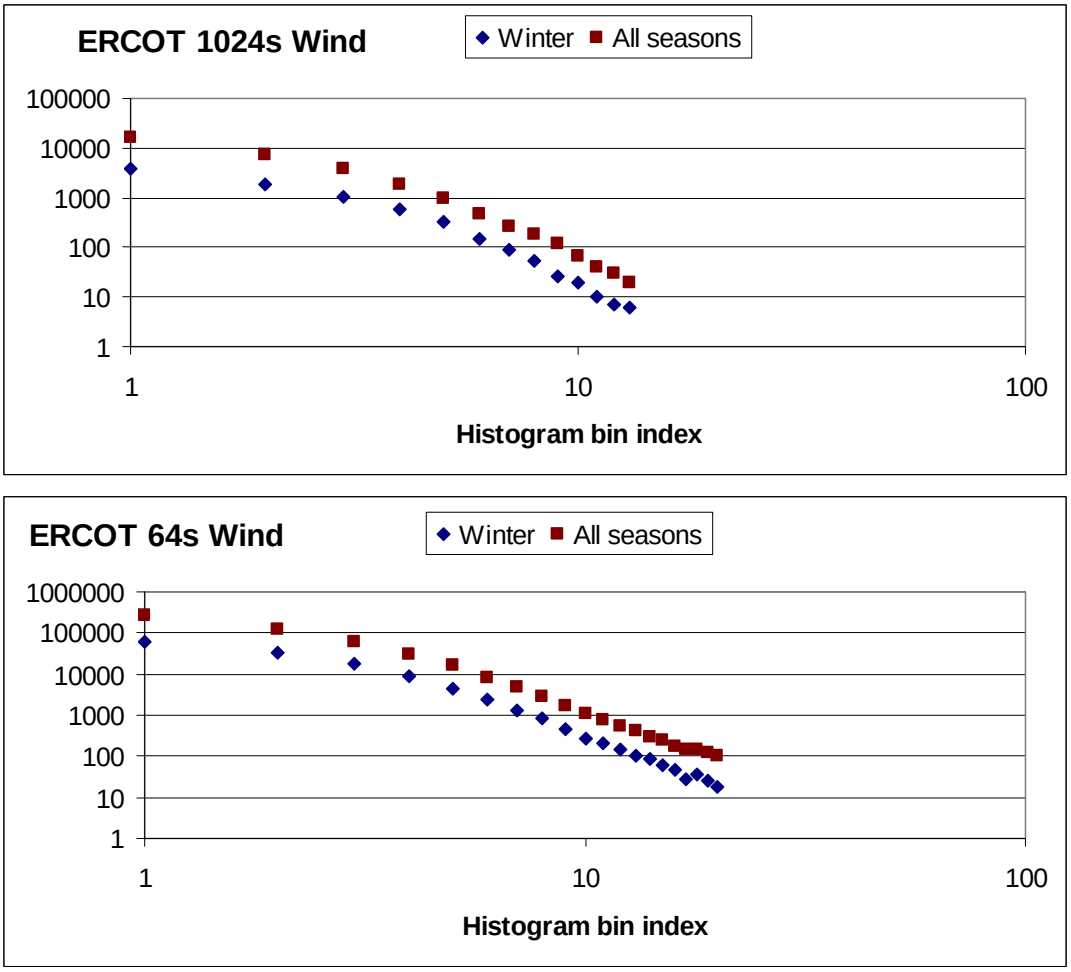


Figure 4-2. Bin count *vs* bin index on a log-log plot for ERCOT.

One advantage of developing an empirical estimate of the distribution shape is that it can be used to estimate the potential occurrence of extreme events from relatively short time series (i.e. a few years). If it is true that the wind power deltas follow a power law distribution in the tails, a simple expression can be derived for the probability of large events. Let p(x) be the probability density for fluctuations of size x: If p(x) follows a power law for x > a, with exponent $-\beta$ ($\beta>0$), then the probability density function can be written[4] (Clauset *et al*. 2007)

$$p(x) = [\,(\beta-1)/a\,]^{*}\,(\,x/a\,)^{-\beta}, \quad x > a, \text{ and } \beta > 1.$$

To estimate the probability of seeing a value of x larger than L > a, the probability density function can be integrated to get

$$Q_L = \int_L p(x)\,dx = (\,a/L\,)^{\beta-1}$$

[4] For simplicity this formula assumes that p(x) is zero for x < a; if this is not the case the normalization constant must be modified accordingly.

Quantitative estimation of Q_L requires the determination of the parameters a and β, which will depend on the details of the system under consideration. Longer data sets are required to make reliable estimations of distribution shape parameters (Clauset *et al*. 2007).

4.3 Concluding Remarks

In this study we have provided a new view of wind and load time series data, making no assumptions about how the data should behave. Using the wavelet transform approach we have converted the original time data to a collection of time series of the smoothed data and the complementary differences for a range of time scales from a few seconds to about one hour. This provides a simple way to examine the statistical behavior of the data as a function of time scale, while avoiding the introduction of algebraic dependence in the smoothed and differenced signals.

Modeling studies for large rates of wind penetration must rely on simulations of wind power output. Typically, meteorological models are used to produce wind speeds and these are converted to wind power output based on several engineering models of turbines. Weather models tend to produce overly smooth wind speeds (WWSIS 2008), so additional variability must be added in to represent fluctuations at the minute time scale. These statistical approaches tend to assume that short-term fluctuations are distributed according to a Gaussian curve. Our examination of the data shows clearly that this is a very poor approximation to the real behavior. This is likely to impact the results of modeling studies for three reasons.

1. Relative to an exponential, a Gaussian concentrates much more mass near the center and will consequently over-estimate the probability of small to moderate deviations. This in turn may lead to an over-estimation of the additional reserve requirement, and associated costs, for high penetration of wind.

2. A Gaussian decays very rapidly in the tails, whereas the data suggest that the real distribution has relatively long tails. Using a Gaussian to model the distribution provides no useful information about the actual likelihood of large changes in wind power output.

3. Correlations between wind and load do exist at time scales of one to several minutes. These correlations are not captured by adding random variability to the longer time-scale wind model output. The correlation patterns are complex and depend on the season.

These factors may explain why the results of the ERCOT study (GE 2008) are different from what we see in the data – in particular that study finds that the variability of net load compared to load increases with time scale, which we do not see here. For time series of one year, the empirical data show that on at least 99% of the wind power deltas lie within 3-sigma of the center; hence, the "3-sigma" rule based on the empirical standard deviation is adequate for current operations.

As wind capacity increases, it becomes more important to have a better representation of the real likelihood of large changes in wind power. We have presented an argument as to why the physical processes underlying wind power fluctuations should produce a power law distribution in the tails, and why this distribution shape should become independent of system size for sufficiently large systems. We note that the tail behavior for the NYISO and PJM systems,

described in the appendices, does not look much like a power law, but this is not surprising given the small size of these wind systems. The fact that the functional form of the distribution becomes constant does not mean that statistical distributions for the larger system can be constructed just by multiplying the deltas by a constant scale factor (cf. Wan 2005). For any given system size, the distribution parameters (the values of a and β from section 4.2) must be computed directly from the data. To evaluate the power-law hypothesis quantitatively, more data are needed. We estimate that at least 3 years of time series data for a collection of widely distributed individual wind plants would be required, with a minimum of 50 sites, distributed over an area of several hundred square miles. By sampling a random collection of sites within the full set, one could construct a set of sub-systems of varying sizes. Applying the methods described here to each subsystem would allow validation of our hypotheses on the distribution shape. Such an analysis could also be used to quantify the relationship between the distribution parameters (a and β) and the system size in MW, and determine whether this relationship is sufficiently regular to be extrapolated to larger systems. Validation of the power-law shape, and estimation of the shape parameters, would be extremely useful in improving the models used to simulate wind power in large-scale integration studies.

It appears generically that, at certain times of day, wind power can lead to substantially higher ramps that can be more difficult to manage. An important operational question is whether one can develop a method to create short-term, real-time forecasts of ramp size and duration (Ela et al. 2009, GE Energy 2008, CAISO 2007). Since a short-term forecast will be based on some combination of the same-day wind conditions and general statistics, it is inherently a problem of evaluating correlations between the wind deltas at different time scales. To correctly extract these correlations from data, an approach like the wavelet method is needed, to ensure that the deltas at different time scales do not contain redundant information.

References

Apt, Jay. 2007. The spectrum of power from wind turbines. *Journal of Power Sources* 169, no. 2 (June 20): 369-374. doi:10.1016/j.jpowsour.2007.02.077.

AWEA 2010. American Wind Energy Association Projects Database (Texas). http://www.awea.org/projects/Projects.aspx?s=Texas. Accessed March 15 2010. These data are also summarized in the AWEA quarterly reports.

Burrus, C. S., R. A. Gopinath & H. Guo 1998. Introduction to wavelets and wavelet transforms: a primer. Prentice Hall.

CAISO 2007. California Independent System Operator. Integration of Renewable Resources. November 2007.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J Newman. 2007. Power-law distributions in empirical data. *0706.1062* (June 7). doi:doi:10.1137/070710111. http://arxiv.org/abs/0706.1062.

Ela, Erik and Jason Kemper. 2009. Wind Plant Ramping Behavior. National Renewable Energy Lab Report NREL/TP- 550-46938. December 2009.

Enernex. 2006. Minnesota statewide wind integration study. November 2006. http://www.enernex.com/projects/project_wind_minesota.htm.

GE Energy 2008. Analysis of Wind Generation Impact on ERCOT Ancillary Services Requirements, Final Report. March 28, 2008.

Wan, Y., and J. R Liao. 2006. Analyses of Wind Energy Impact on WFEC System Operations. National Renewable Energy Lab Report NREL/JA- 500-39583. May 2006.

Wan, Yih-Huei. 2005. Primer on Wind Power for Utility Applications. National Renewable Energy Lab Report NREL/TP- 500-36230. December 2005.

Wan, Yih-Huei. 2004. Wind Power Plant Behaviors: Analyses of Long-Term Wind Power Data. National Renewable Energy Lab Report NREL/TP- 500-36551. September 2004.

Woyte, Achim, Ronnie Belmans, and Johan Nijs. 2007. Fluctuations in instantaneous clearness index: Analysis and statistics. *Solar Energy* 81, no. 2 (February): 195-206. doi:10.1016/j.solener.2006.03.001.

WWSIS 2008. Western Wind and Solar Integration Study. Development of Regional Wind Resource and Wind Plant Output Datasets Final Report October 15, 2007.

Appendix A. PJM Wind Data

The wind power data for the PJM control area cover the period January 1 2008 through December 31 2008 at five-second resolution. The time series data were separated into four 96-day seasons as for ERCOT. For PJM the months (December, January and February) that comprise the winter season are not contiguous. To get around this problem, the December data were relabeled as 2007, and the three months treated as if they were contiguous. For the statistical analysis, in principle it makes no difference what year the data come from, so this approach should not have any significant impact on the metrics used here. It could however introduce a large jump in the deltas due to the discontinuity between December 31 and January 1. Our method of screening for outliers (defined as 2 times the 99th percentile value) would remove the impact of this large jump on the correlation coefficients. To confirm that the use of non-contiguous data does not lead to any qualitative change in the results, we also examined the statistical metrics for a 64-day time series based on the months January and February, and for a 96-day time series based on January, February and March, All these data sets show qualitatively the same behavior; quantitative differences exist of the order of a few percent.

The PJM data were provided as individual time series for 26 wind power plants. Some of these plants operated during only part of the data year. As we use the absolute values of the deltas in this analysis, to be consistent the data set should describe wind output for a system of constant capacity over the data year. Hence, only the subset of plants that were operating over the entire year is included in the analysis. The on-line dates for the various wind farms were not provided, so we used the monthly maximum values of output to determine whether a plant was fully operational during that month. If, for each month in the data year, the monthly maxima differed from one other by less than 10 percent, the plant was deemed to be fully operational for the whole year. Based on this approach a subset of 16 of the wind farms were used to construct the seasonal data for this analysis. Using the maximum output value for each farm as an estimate of its capacity, for the full set of wind farms in the PJM data the capacity is approximately 2620 MW, and for the subset used in this analysis the capacity is 1060 MW. The average output for the analysis subset was 420 MW for winter and 181 MW in summer.

Table A-1 shows the winter and summer statistics for the wind deltas, including the RMS values, and the average, standard deviation 99th percentile and maxima for the positive and negative deltas separately. The data reproduce many of the features seen in the ERCOT data: for short times scales the RMS values (or the average plus standard deviation for deltas of a given sign) are relatively independent of time scale, while at longer time scales they grow due to accumulation of changes with the same sign. The 99th percentile values are significantly smaller than the maxima, indicating potentially long tails in the distributions.

Table A- 1. Statistics for PJM wind deltas, summer and winter.

	Winter (DJF)								
Tsec	rmsdiff	max+	p99+	av+	sd+	max-	p99-	av-	sd-
10	2.5	120.0	4.5	1.1	3.0	88.5	4.4	1.0	1.5
20	3.1	90.3	6.2	1.6	2.5	119.3	6.2	1.6	2.9
40	3.7	107.4	7.6	2.0	3.2	139.7	7.5	2.0	3.1
80	4.0	124.9	9.3	2.3	3.3	114.6	9.3	2.3	3.2
160	4.9	116.3	13.2	2.9	3.9	206.3	13.0	2.9	4.1
320	6.7	120.2	21.0	4.2	5.2	120.9	21.0	4.2	5.3
640	10.7	123.4	36.9	7.0	8.1	133.5	35.4	6.9	8.2
1280	17.5	171.2	67.4	11.7	13.4	173.2	62.4	11.2	13.0
2560	26.0	139.2	99.5	18.9	19.6	139.5	77.9	17.7	17.3
	Summer (JJA)								
Tsec	rmsdiff	max+	p99+	av+	sd+	max-	p99-	av-	sd-
10	0.9	42.1	3.1	0.7	0.7	40.8	3.0	0.7	0.7
20	1.3	40.4	4.3	1.0	0.9	36.7	4.3	1.0	0.9
40	1.8	17.4	5.9	1.3	1.3	38.1	5.7	1.2	1.3
80	2.4	81.1	8.3	1.6	1.8	52.2	7.9	1.6	1.7
160	3.5	40.3	13.0	2.3	2.7	47.5	12.4	2.3	2.6
320	5.9	77.0	22.9	3.7	4.8	72.6	21.2	3.7	4.5
640	10.5	104.7	42.6	6.4	8.5	173.7	40.0	6.3	8.3
1280	17.6	161.2	68.6	10.6	13.9	177.3	68.0	10.7	14.1
2560	26.8	188.4	101.4	16.5	19.7	221.2	102.5	17.7	21.6

Distributions of the wind power deltas for three short and three long time scales are presented in figures A-1 and A-2. These figures follow the conventions used for the ERCOT data; the vertical axis is logarithmic. The bin size for the short time scales is 1.08 MW and 5.74 MW for long time scales. The distribution shapes are less sharply peaked than for ERCOT. At long time scales, the histogram shape looks exponential for bin indices up to about +/- 7, and broadens out in the tails. The shorter time-scale data, particularly the 40-second data, show some rounding at the very center of the distribution, which would be consistent with a somewhat more Gaussian shape near k=0. However this is a minor feature and the distribution shape is still predominantly exponential.

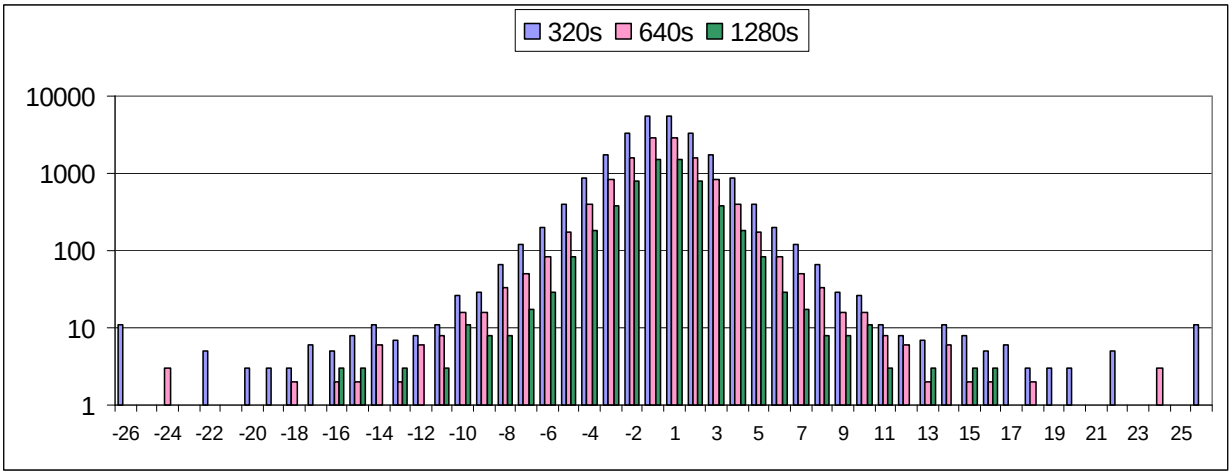


Figure A- 1. Frequency distribution of PJM wind power deltas at long time scales

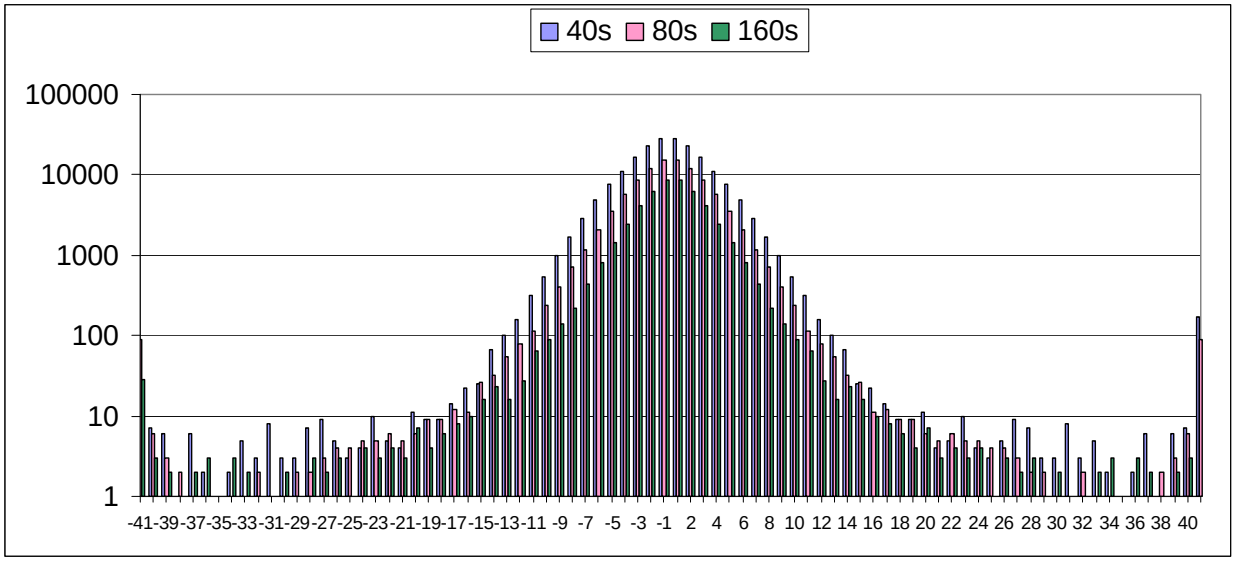


Figure A- 2. Frequency distribution for PJM wind power deltas at short time scales

Appendix B. NYISO Wind Data

The wind power data for the NYISO area cover the period September 1 2008 through April 30 2009 at six-second resolution. The data were used to construct two 96-day seasons, fall and winter. As with the PJM data, time series for a set of 15 wind power plants were provided separately. The procedure described above for PJM was used to identify those sites that were operating continuously during the months September through February. Based on values of the monthly maximum for each plant, our capacity estimate for the full NYISO data set (as of April 2009) is about 1350 MW. After dropping those sites that became operational during the data period, the resulting set of 11 sites corresponds to about 700 MW of capacity. For this subset, the average output is 256 MW in winter and 136 MW in fall. Statistical measures for this data are given in Table B-1. In Figures B-1 and B-2 we show the long and short time-scale histograms for the wind deltas. The bin sizes are 1.0 MW for the short time scale and 5.5 MW for the long time-scale. Once again the distributions show a pronounced exponential shape; in this case there is little evidence of long tails.

Table B- 1. Statistics for NYISO wind deltas, fall and winter

	Winter (DJF)								
Tsec	rmsdiff	max+	p99+	av+	sd+	max-	p99-	av-	sd-
12	1.7	31.7	6.0	1.1	1.3	65.3	5.9	1.1	1.3
24	2.2	36.5	7.1	1.5	1.6	58.5	7.0	1.5	1.6
48	2.7	69.9	8.3	1.9	1.9	63.7	8.4	1.9	1.9
96	3.1	100.8	9.3	2.2	2.2	52.2	9.5	2.1	2.2
192	3.7	114.5	11.5	2.6	2.7	55.7	11.3	2.5	2.6
384	5.3	151.7	17.1	3.6	4.0	56.9	17.0	3.6	3.8
768	8.4	82.7	26.5	5.9	6.2	77.7	27.0	5.8	5.9
1536	15.4	245.6	49.7	10.0	11.8	288.1	43.5	9.8	11.8
3072	24.6	383.9	80.9	16.4	20.7	173.2	68.7	15.6	16.6
	Fall (SON)								
12	1.2	76.8	4.3	0.7	1.0	246.1	4.3	0.7	1.1
24	1.5	38.5	5.5	1.0	1.2	177.0	5.4	1.0	1.3
48	2.0	29.4	6.7	1.2	1.5	211.0	6.8	1.2	1.7
96	2.4	33.6	8.0	1.5	1.8	197.2	7.9	1.5	2.1
192	3.1	40.1	10.4	1.9	2.3	204.1	9.8	1.9	2.8
384	4.8	62.0	15.6	2.8	3.5	199.4	15.2	2.8	4.4
768	7.5	79.5	27.6	4.7	5.8	138.0	24.7	4.5	6.1
1536	15.6	376.1	46.2	8.3	14.7	370.1	42.9	7.7	12.6
3072	20.0	181.7	62.2	12.1	14.9	191.8	76.1	12.6	16.8

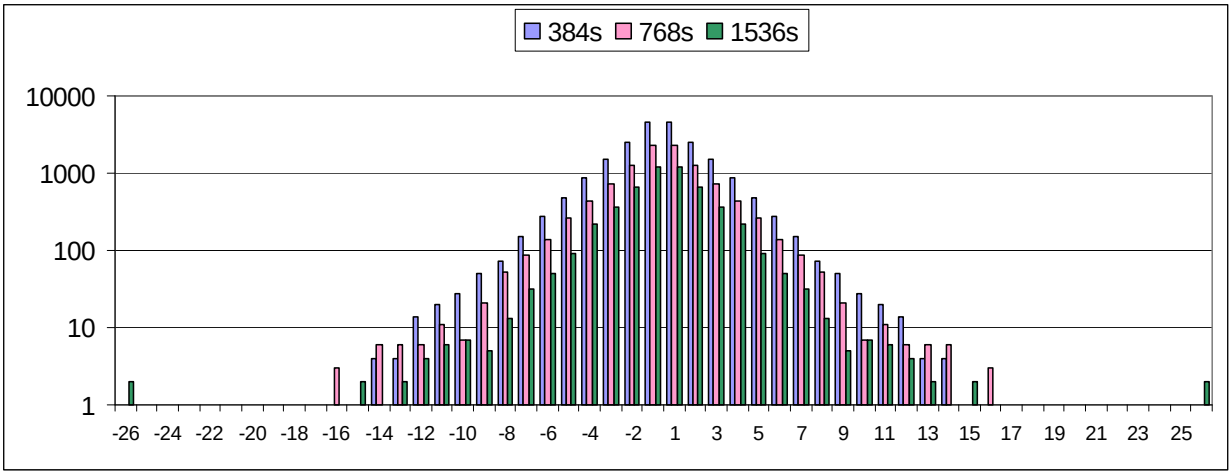


Figure B- 1. Frequency distribution of NYISO wind power deltas at short time scales

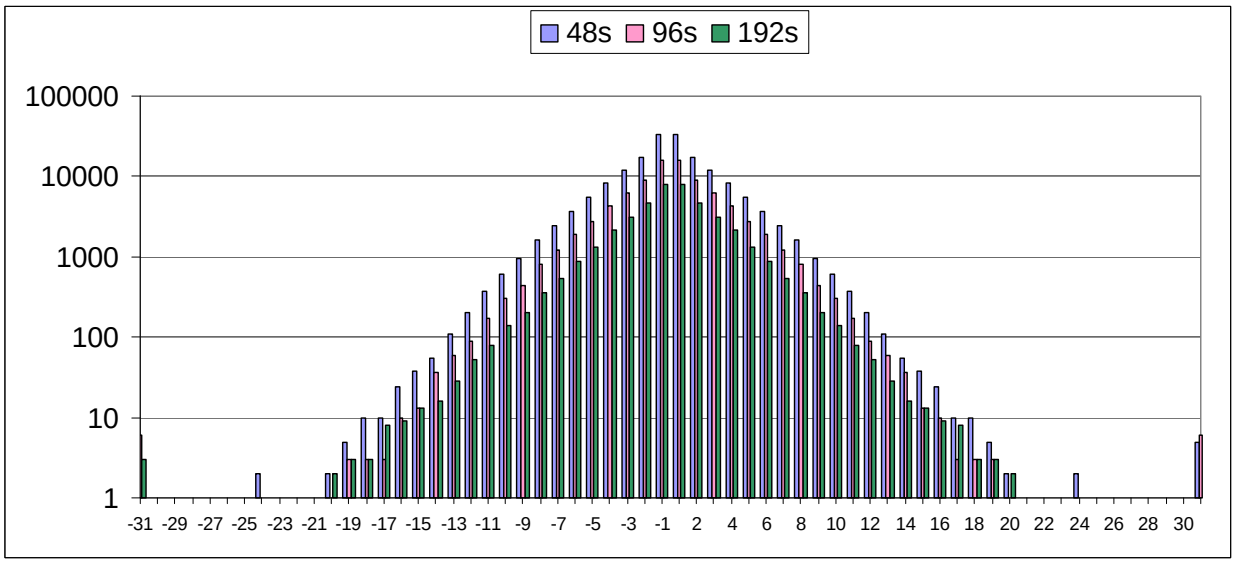


Figure B- 2. Frequency distribution of NYISO wind power deltas at short time scales