NREL
NATIONAL RENEWABLE ENERGY LABORATORY

Wind Data Inputs for Regional Wind Integration Studies

Preprint

D. Lew, K. Orwig, and Y. Wan
National Renewable Energy Laboratory

C. Alonge, M. Brower, and J. Frank
AWS Truepower

L. Freeman
GE Energy

C. Potter
3TIER Group

To be presented at the Power & Energy Society Annual Meeting
Detroit, Michigan
July 24-29, 2011

Conference Paper
NREL/CP-5500-50636
March 2011

Contract No. DE-AC36-08GO28308

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information

P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Wind Data Inputs for Regional Wind Integration Studies

Debra Lew, *Member, IEEE*, Charles Alonge, Michael Brower, Jaclyn Frank, Lavelle Freeman, *Member, IEEE*, Kirsten Orwig, Cameron Potter, *Member, IEEE*, and Yih-Huei Wan, *Senior Member, IEEE*

Abstract—**Wind integration studies are conducted routinely to examine the operational impacts of wind on the power system. Wind plant power outputs and forecasts are needed as inputs to these studies and this data is often synthesized by a variety of methods. This paper examines the methodologies used to create these datasets, the pitfalls that may be encountered, and the tradeoffs between different methodological approaches.**

Index Terms—**Wind power modeling, wind speed, mesoscale modeling, forecast error, wind forecasting, and wind integration.**

I. Introduction

Utilities in the United States have seen high interest in wind development, as evidenced by the fact that wind leads all other fuels and technologies in interconnection requests [1] and that wind competes with solar as the nation's fastest growing energy source [2]. As a result of this interest, utilities, Regional Transmission Organizations (RTOs), and Independent System Operators (ISOs) are increasingly undertaking integration studies to examine the operational impacts of growing penetrations of wind power on their grids.

Many wind integration studies have been conducted to date [3]-[7], including two new regional studies looking across much of the eastern and western interconnections [8],[9]. These studies typically include hourly production cost modeling of at least one year of the power system with various levels of wind power. Each wind plant is modeled, at least on an hourly basis, and provided as inputs to the production cost model, along with hourly load data. In addition to wind output data, wind forecast data is also typically used to more accurately model how a utility or RTO/ISO might commit or dispatch units based on a forecast.

While existing wind plants can be represented with historical output data, hypothetical future plants cannot. Onsite measurements for wind speed can be converted to power and used as data inputs. However, some of these studies, for example, the Eastern Wind Integration and Transmission Study (EWITS) and the Western Wind and Solar Integration Study (WWSIS), examine up to 30% wind penetration, which requires modeling of 330 GW and 75 GW of wind plants, respectively. Even those studies with modest penetration levels typically don't have onsite measurement data to represent the hypothetical plants.

Many studies turn to modeled wind data to represent hypothetical future plants. One methodology seeks to scale and/or time-shift actual measurements to represent nearby hypothetical plants [10]. These may provide realistic outputs in regions where there is a predominant wind direction and where wind plants do not cover a large region. However, for wide geographical regions with more complex terrain, and in order to capture a variety of weather patterns, more sophisticated methodologies may be needed to capture correlations in both space and time across that region. A more commonly used methodology involves physics-based modeling of the atmosphere and the weather patterns over time and on a three-dimensional spatial grid. This Numerical Weather Prediction (NWP) model can then be sampled in time and space to yield hypothetical wind speed measurements that can then be converted to wind plant output. These mesoscale models are typically run at a high horizontal spatial resolution (less than 10 km).

In this paper, historical observations of actual plant output are called *observations*, modeled plant output is called *modeled* or *actuals* (to distinguish them from forecasts), and modeled forecasts of plant output are called *forecasts*. The difference between the forecasts and the actuals are the *forecast errors*. A good wind dataset for use in integration studies will have the following attributes:

- Appropriate characterization of variability on the 1-minute, 10-minute, hourly, daily, and seasonal time scales. Accurate characterization of variability and forecast error is the most important criterion for a dataset. Integration studies are conducted to determine the impact of wind variability and uncertainty on various reserve requirements.
- Appropriate characterization of forecast errors. Forecast errors tend to be higher during wind ramp events. Capturing the appropriate magnitude of forecast error over a single plant, state, or region is important. Forecast error distributions tend to be wider than Gaussian distributions.
- Both modeled and forecasted data should display reasonable spatial correlation, both within a single plant and between plants in a region.
- Both modeled and forecasted data should display reasonable temporal correlation over short, medium, and

This work was supported by the U.S. Department of Energy Wind Program.

D. Lew, K. Orwig and Y-H. Wan are with the National Renewable Energy Laboratory, Golden, CO 80401 USA (email: debra.lew@nrel.gov, kirsten.orwig@nrel.gov, yih-huei.wan@nrel.gov).

C. Alonge, M. Brower, and J. Frank are with AWS Truepower, Albany, NY 12205 USA (email: calonge@awstruepower.com, mbrower@awstruepower.com, jfrank@awstruepower.com).

L. Freeman is with GE Energy, Schenectady, NY 12345 USA (email: Lavelle.freeman@ge.com).

C. Potter is with the 3TIER Group, Seattle, WA 98121 USA (email: cpotter@3tier.com).

longer term timescales. Forecast errors should exhibit appropriate autocorrelation characteristics.

- Accuracy in resource quality as defined by the capacity factor of the site is deemed a less important criterion for datasets. If the capacity factor is low, then additional sites will be needed to reach a target penetration level. This may impact the overall variability since more sites will be aggregated.

II. Results

This paper reviews methodologies used to create regional wind datasets for integration studies, with a focus on the EWITS and WWSIS datasets. A recent Hawaii wind dataset, which built upon the lessons learned in the EWITS and WWSIS datasets and which required higher temporal and spatial resolution and a higher degree of accuracy, is also discussed. The paper investigates the tradeoffs between different approaches as well as some of the pitfalls encountered in creation of these datasets.

A. Spatial Correlation and Consistency

Various mesoscale models are employed to simulate the weather. For example, the WWSIS used the Weather Research and Forecasting (WRF) model and EWITS used the Mesoscale Atmospheric Simulation System (MASS). In both models, a key input was the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Global Reanalysis (NNGR) dataset, which represents the overall state of the atmosphere based on surface and upper air observations.

Modeling a large region at high resolution can quickly exhaust computational capacity limits. EWITS and WWSIS required modeling half of the continental United States at a 2-km, 10-minute resolution over three years. In WWSIS, this meant modeling 1.2 million grid cells. The region was divided into four overlapping domains (see Figure 1), partly because the northwest domain had already been modeled under a separate project and partly to satisfy computational limitations [11]. To eliminate spatial seams resulting from this approach, data from overlapping regions were blended.

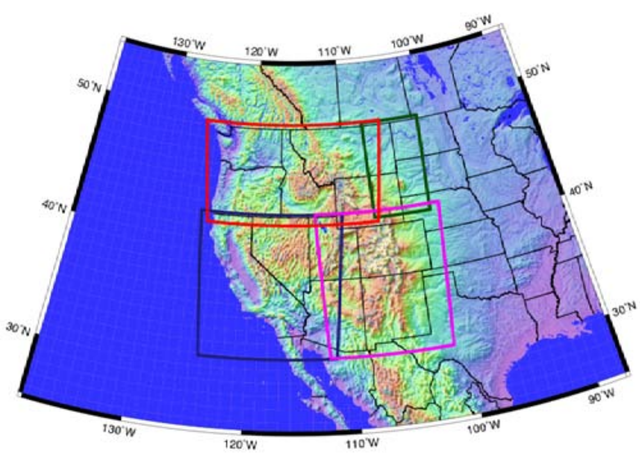

Fig. 1. Four modeling domains used in the WWSIS [11].

EWITS used a nested grid approach, using a 30-km resolution parent grid shown by the large box in Figure 2 and three smaller grids of 8 km and then finally 2-km resolution child grids shown by the red boxes [12]. In this way, grids of a manageable size could be run at a high resolution while maintaining the correlation of the parent grid.

At this time, detailed analysis of the spatial correlation and consistency between these two approaches has not been undertaken. However, it is noted that the blending process used in modeling individual domains is an extra computational step and that such steps are opportunities for additional errors (especially on datasets of this size).

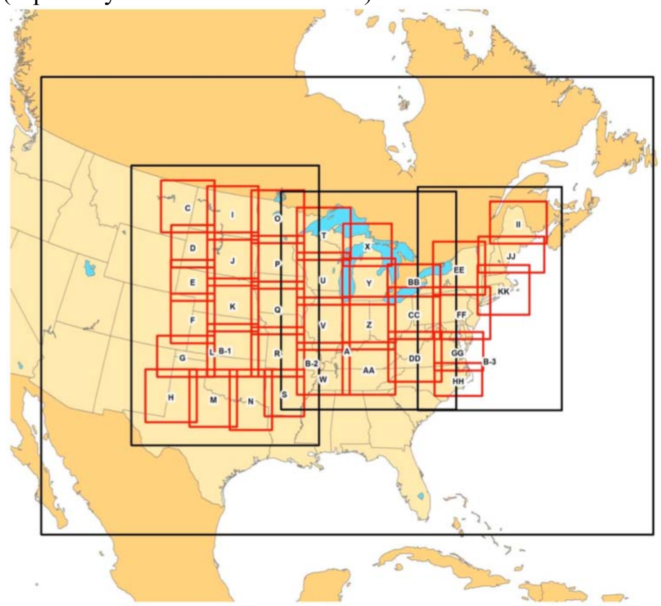

Fig. 2. Nested grids for the EWITS mesoscale model [12].

Mesoscale modeled data can display less spatial diversity than observations. For example, the WWSIS dataset was compared to six operating wind plants in Texas, totaling nearly 540 MW. The modeled data shows a coincident peak of 99.8% of installed capacity for the years 2004-2006, while the observations show a coincident peak of 86.5%.

B. Temporal Correlation and Consistency

The mesoscale models are run for specific time intervals and then restarted using NNGR input data. The length of this time interval must strike a balance between allowing the mesoscale model to establish and capture mesoscale flows and patterns, and aligning the model with the higher level NNGR analysis data. Typically, the mesoscale model for the next time interval is spun-up for about 12 hours before data from that model is used, to allow the model to establish the mesoscale circulations. Model runs in WWSIS and EWITS were 3 and 15-16 days, respectively.

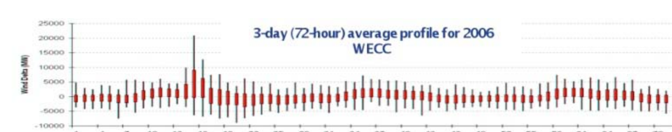

Fig. 3. Hourly changes in wind power output for wind sites in WWSIS scenarios across the Western Electricity Coordinating Council/WECC for 2006. Data has been parsed into 3-day intervals. The long rectangles show the one sigma variability and the whiskers show the maxima and minima [8].

In WWSIS, the five hours around the model restart were blended to minimize the impact of the temporal seam. In addition, one-hour deltas (hour-to-hour ramps) were checked to ensure that no hourly ramp was outside of plausible limits. Despite this, the hours around the model restart displayed statistically significantly higher variability than other hours. While this did not impact the hourly production simulation analysis, it did impact the statistical analysis of WWSIS. As a result, every third day was removed from the dataset in the statistical analysis. Figure 3 shows the average profiles of one-hour deltas for the wind plant output data for all of the Western Electricity Coordinating Council (WECC) when the data has been parsed into three-day intervals. The restarts occurred at 00 GMT, or 4 pm local time on Day 1.

Another potential entry for temporal issues is the assimilation of rawinsonde data. In EWITS, rawinsonde and surface observation data were assimilated every 12 hours (at 00 and 12 GMT) to align the model with observations. Figure 4, top, shows the abrupt jumps that were found on many of the days. Figure 4, bottom, shows the effect of replacing the affected hours with synthesized data.

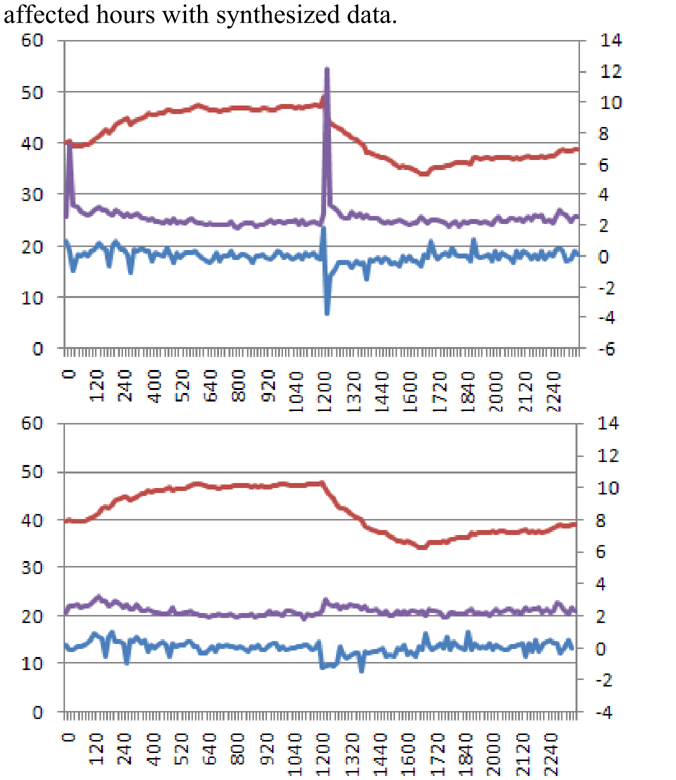

Fig. 4. Jumps in wind power output due to abrupt assimilations of observations every 12 hours, showing before (top) and after (bottom) the fix [12]. The left Y-axis shows wind plant output in megawatts/MW, the right Y-axis shows changes in wind output in MW, and the X-axis shows time in GMT. The red curve is the mean output (left axis), the purple curve is the absolute value of the change in output from one 10-minute record to the next (right axis), and the blue curve is the change in output (right axis).

As a follow-up to EWITS, the same methodology was used to develop a higher resolution dataset for Hawaii [13]. While the fix shown in Figure 4 resolved the discontinuity in the wind power output, a discontinuity remained in the first derivative. The Hawaii power system is much smaller and the penetration levels studied much higher than in EWITS, so this discontinuity in the first derivative needed to be eliminated for the integration analysis. The high frequency signal was retained and 10-minute wind speed changes (eliminating the unrealistic jump) from that signal was randomly applied to the wind speeds (see Figure 5).

In the Hawaii dataset, the largest wind event (a significant wind down ramp at three islands) happened to occur at a time both when the model was restarted and when rawinsonde data was assimilated. To determine whether the event was due to the model restart or the observational assimilations, three runs were compared: the original run with the restart and assimilation, the original run without the assimilation of data and a run without the restart but with assimilation of data. The results of these three runs are shown in Figure 6 for one site. They show that the model restart was the culprit for the large ramp event at 00 GMT on Oct. 16, and that the assimilation of rawinsonde data was not a factor.

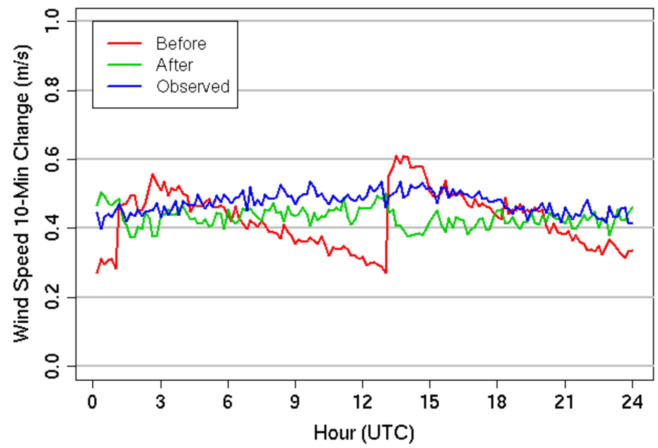

Fig. 5. Ten-minute absolute changes in wind speed at wind plant in Hawaii before adjustment (red), after adjustment (green), and observed (blue) [13].

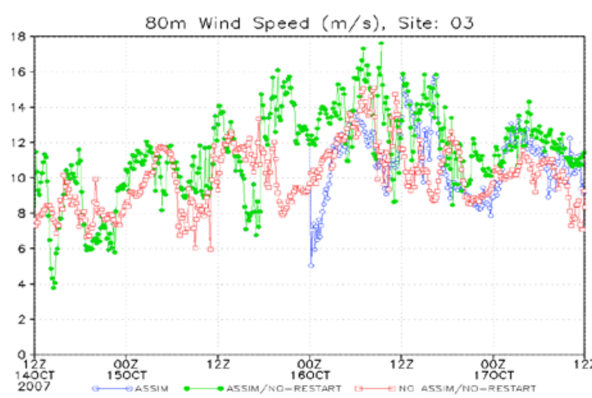

Fig. 6. Comparison of three different model runs: original run with restart and data assimilation (blue), data assimilation with no restart (green), and no assimilation/no restart (red) [13]. Y-axis shows wind speed in m/s and x-axis shows time in GMT. Prior to 00 GMT on 16 Oct, the green trace overlays the blue trace so that it is not easily seen in this graph.

The temporal seam at model restart was judged too difficult to fix. In integration studies, detailed examination of significant events is often conducted to ensure the power system can operate through these events or to determine what mitigation options must be undertaken so the system can operate through these events. It was recommended that the selection process for significant events ignore the model restarts (every 15-16 days).

It is important that integration analysts have a detailed understanding of how their wind data was generated so that they can identify and resolve periods of increased variability that may be artifacts of the modeling process and not truly representative of wind plant behavior.

C. Wind Speed Validation

To check the accuracy of the wind speed modeling, detailed validation analyses are conducted to compare modeled data to anemometer observations during the same time period. It is important to ensure that the same data used in developing the modeled data is *not* used in validating the modeled data.

Fig. 7. Locations of publicly available tower data that were used to validate wind speeds in WWSIS [11].

TABLE I
SUMMARY OF RESULTS FROM WIND SPEED VALIDATION OF WWSIS DATASET AGAINST 28 PUBLICLY AVAILABLE TOWER OBSERVATIONS [11]

Tower	State	Height [m]	Observed wind speed [m/s]	Modeled wind speed [m/s]
3001	AZ	30	5.47	5.95
3002	AZ	30	6.40	6.23
3003	AZ	30	5.43	5.47
3006	AZ	30	4.80	4.47
4402	CA	47	4.24	4.40
4403	CA	61	5.89	4.55
6001	CO	44	7.37	6.47
6008	CO	20	5.29	4.84
6009	CO	20	4.30	4.36
6013	CO	80	4.80	6.02
6029	CO	20	5.87	5.27
6039	CO	50	7.71	7.56
12111	ID	60	6.49	6.03
12131	ID	20	5.90	5.01
12439	ID	20	5.51	4.25
12500	ID	20	6.29	4.39
12505	ID	20	5.45	3.79
26007	MT	30	7.39	6.24
26010	MT	40	7.62	6.59
28001	NV	50	4.62	4.82
28002	NV	50	4.46	5.50
28003	NV	50	5.46	4.97
31010	NM	70	8.54	8.07
31011	NM	39	7.02	6.96
34018	ND	40	7.07	6.98
44003	UT	20	3.98	3.76
44022	UT	20	4.90	4.48
44999	UT	50	6.42	6.31

WWSIS used a MOS (model output statistics) correction to adjust the raw mesoscale model output. The correction was used with the Rapid Update Cycle (RUC) dataset produced by the National Oceanic and Atmospheric Administration (NOAA) and NCEP. The RUC dataset extensively uses 10-m towers which can make this correction less valid in areas of complex terrain, so the correction was weighted.

In WWSIS, wind speed validation was undertaken with 34 publicly available towers and also some proprietary tower data. The locations of 28 of the public towers (Pacific Northwest towers are not shown here) are shown in Figure 7. Note that there is limited publicly available observational data, especially in Wyoming where some of the best resources are located. Proprietary data can help to some degree. Even where there are publicly available observations, many of these are far below hub height (20-50 meters). Periods of record are often a small subset of the three years (2004-2006) modeled. This underscores the need for publicly available wind data.

Summary results from the validation against the 28 publicly available towers shown in Figure 7 are shown in Table I. Accuracy and bias varies, even within a single state. For example, northeast Colorado winds are underestimated while winds just north of Denver are overestimated. This makes it difficult to apply a simple adjustment to further increase the accuracy of the dataset. It is important to note that these are wind speeds, and that wind power is proportional to the cube of the wind speed.

D. Wind Power Conversion

Ultimately for wind integration studies, wind *power* output is needed. Conversion of a single wind speed measurement to wind plant power output is a non-trivial exercise. Wind speeds vary across the 2 km resolution of the grids in EWITS/WWSIS and across wind plant layouts, which themselves vary from plant to plant. Output of a single wind turbine is stochastic, as is the output of a wind plant.

In WWSIS, a modified version of 3TIER's SCORE (Statistical Correction to Output from a Record Extension) process was used [14]. SCORE creates probability density functions (PDF) from actual wind turbine output and samples from those PDFs to generate hypothetical wind turbine output. SCORE-lite simplified the process for the large number of sites in WWSIS; instead of applying the PDFs to each wind turbine, the PDFs were applied to each grid cell which contained 10 wind turbines. The Vestas V90 3-MW turbine was modeled with a hub height of 100 m, based on observed PDFs from wind plants with these turbines.

Mesoscale models tend to produce wind speeds that are excessively smooth. The SCORE process adds back in some variability so that the wind plant output is more realistic.

In EWITS, AWS Truepower used their SynOutput procedure. AWS Truepower's 20-km historical dataset for 1997-2007 and 200-m wind maps were used to scale the mean wind speeds for each site. Ten validation towers were used to adjust the raw data, depending on which tower was most representative of each particular site. An annual adjustment was made. Three synthetic power curves were used for each IEC class with the IEC 1 and 2 turbines having a hub height of 80 m, and the IEC 3 turbine at 100 m.

Corrections are then made to the wind speed output to ensure that the wind plant power output is realistic. This includes adding in the turbulent kinetic energy (TKE) and applying a time filter to mimic spatial averaging across the grid cell. Wake losses, electrical losses, and plant availability were all applied to produce final wind plant outputs.

It should be noted that neither of these methods is attempting to exactly recreate specific plant output over a specific time window. Instead, they attempt to capture the statistical characteristics of the variability of wind plant output over different timescales. Aggregated over a region, wind ramps tend to occur over the tens of minutes and hours timeframe, so these methods may provide suitable accuracy for system studies. However, for small systems with a small number of wind plants, this data may need further evaluation to determine applicability in system studies.

E. Wind Ramp Validation

The most important validation of a wind dataset for use in integration studies is of the variability of wind plant output. Integration studies examine the impacts of increased variability due to wind on the power system and what may be needed to maintain reliability of the power system. Capturing this variability correctly is important to ensure reliability without overinvestment in unnecessary mitigation options.

Figure 8 shows the comparison of modeled and observed wind plant output ramps for a wind plant in Hawaii.

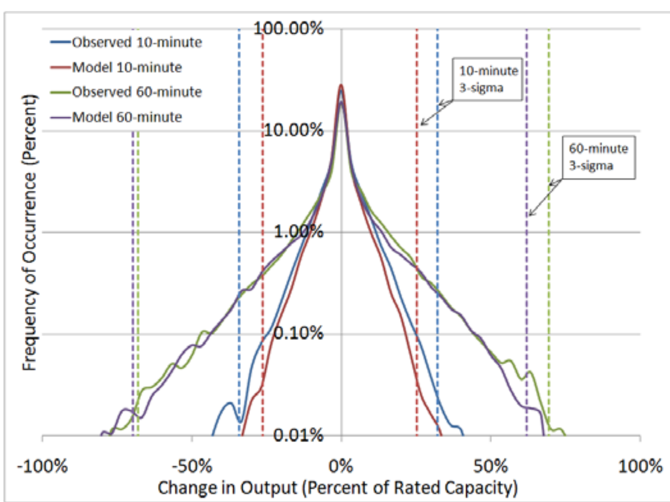

Fig. 8. Comparison of modeled and observed 10-minute and 60-minute wind plant output ramps for a Hawaiian wind plant [13].

Figure 9 shows the comparison of 10-minute wind plant output ramps for a region in the western U.S. It should be noted that when validating ramping behavior at specific wind plants, significantly different results were found, depending on which sites were selected to model the plant even though all sites were very close to each other.

F. Issues with Wind Plant Output Observations

Even observed wind plant output is not free from problems. Wind plants are curtailed; there may be higher than normal outages in the period of record; there may be ramp rate limitations, etc. While validating the wind modeling in the Hawaii dataset, it was found that the modeled wind speeds closely matched the observed wind speeds, but that the wind power outputs were not well-matched (see Figure 10). The difference between the modeled and observed wind plant outputs resulted from a frequent curtailment of the wind plant at night.

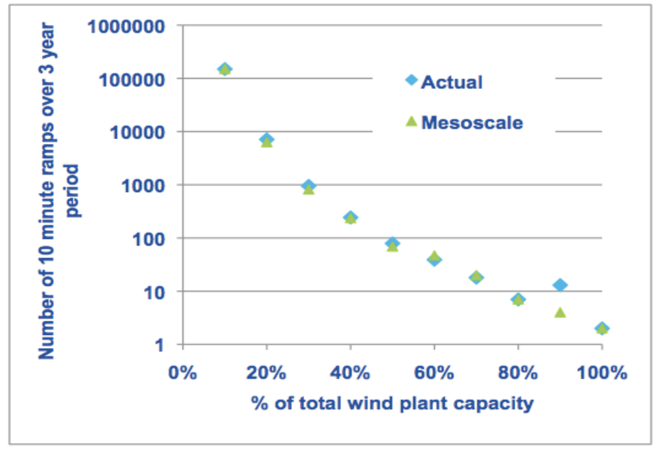

Fig. 9. Comparison of modeled and observed 10-minute wind plant output ramps for a wind plant in the western U.S. [8]

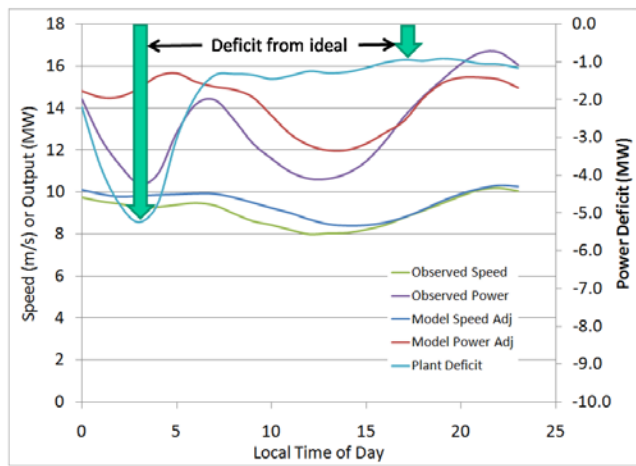

Fig. 10. Observed and modeled power output and wind speed at a Hawaiian wind plant [13].

The Hawaii wind dataset required wind plant output at a 2-second interval. Modeling wind speeds at a 2-second level and converting to power would have introduced too many potential sources of error. Therefore, the entire process was conducted in the power domain. Ten-minute wind plant output data was developed using mesoscale modeling as described above. Two-second power output observations were collected from a wind plant in Hawaii. The underlying 10-minute trend was removed from the output so that the 2-second residuals remained. These 2-second residuals were applied to the 10-minute wind plant output data. However, when the ramping histograms were analyzed (see Figure 11 top), it was found that the ramping behavior was asymmetric. This asymmetry was found to be due to an up-ramp rate limit of 2 MW/minute on the wind plant, and no down-ramp rate limit. In order to model a wind plant without ramp-rate limits, the up ramps were assumed to have the same distribution as the down ramps, and the 2-second data was re-run (see Figure 11 bottom).

Power spectral density (PSD) was used to ensure the high frequency modeled data accurately modeled actual wind plant behavior. Using this technique, high frequency harmonics were identified in the spectra which were caused by limiting the sample size of the 2-second data inputs. When this limitation was removed, the improved spectra in Figure 12 resulted. A kink in the spectra at 20 minutes remained, which was due to a fix of the abrupt jumps from the assimilation of observations in the modeling process.

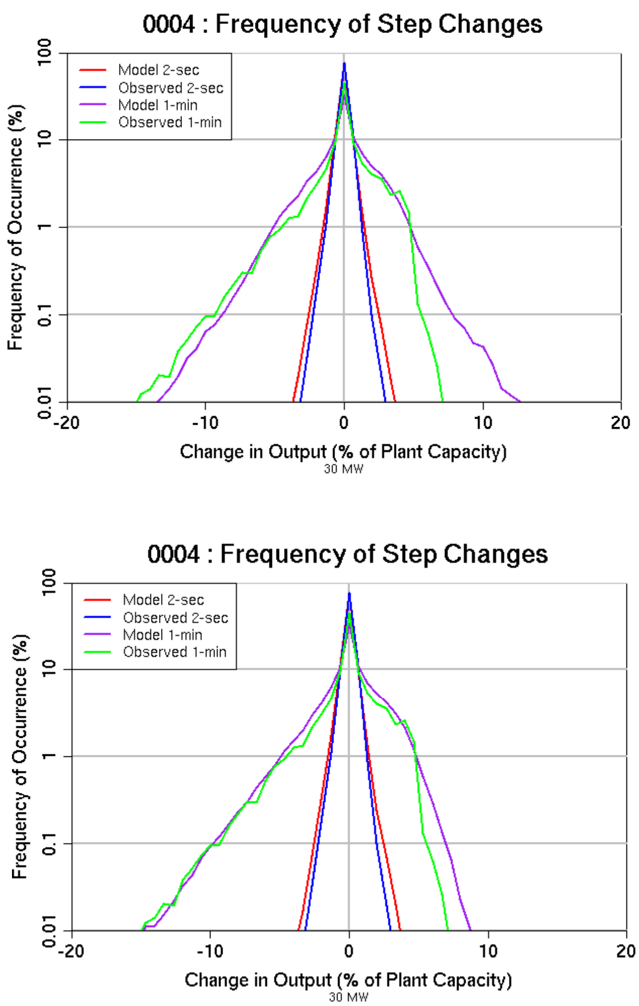

Fig. 11. Frequency distribution of step changes in simulated 1-minute and 2-second output before (top) and after (bottom) the workaround [13].

G. Forecasts

In addition to examining the operational impacts of variability (ramping behavior) of wind plants, integration studies also typically examine the impact of uncertainty (forecast error). There are many ways to develop forecast datasets and it is difficult to determine which is most appropriate because it is typically not clear how forecasts will be developed and utilized. For example, will there be a centralized forecasting provider or will individual wind plant owners develop their own forecasts? This would affect the level of forecast error correlation between sites. What methodology will be used in the future to develop the forecasts? Will they be statistical or NWP-based? All of these would affect forecast error characteristics such as biases, spatial and temporal correlation as well as the overall error.

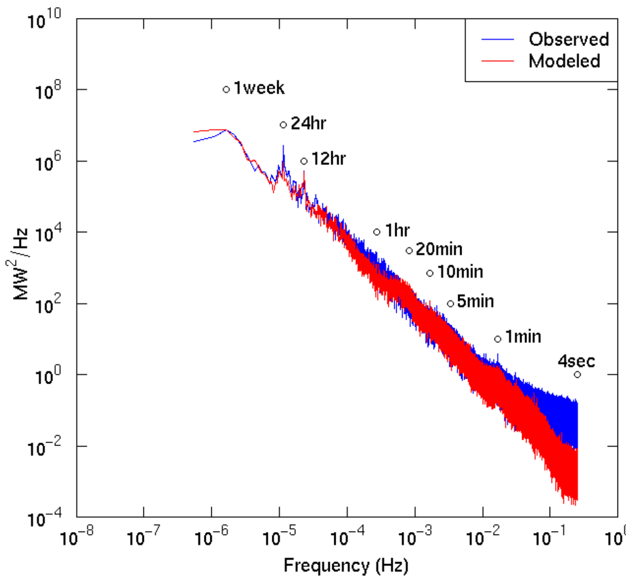

Fig. 12. Power spectral density of observed (blue) and modeled (red) high-frequency data for a wind plant in Hawaii [13].

In addition to these considerations, it is difficult to synthesize both 'actual' wind output and 'forecasted' wind output without generating forecasts that are 'too good' because the same systematic biases were built into both datasets.

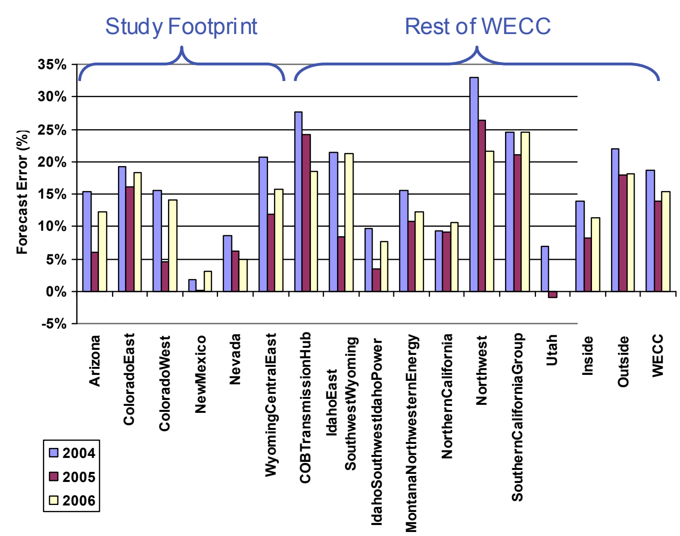

Fig. 13. Forecast error by state/region [8].

In WWSIS, the forecasts were developed by running the same NWP and WRF datasets, but instead of using NNGR as the input dataset, it used the Global Forecast System (GFS), which is the information used to perform state-of-the-art forecasting. This was run at an hourly, 6-km resolution as a single large domain across the western U.S. Day-ahead hourly forecast errors were 12-16% mean absolute error (MAE) on a state basis, but 8% over the study footprint and 7% across all of WECC, which seemed reasonable. However, there was a significant positive bias in the forecasts, with approximately 10% high forecasts in the study footprint and 20% in the rest of WECC (see Figure 13). The forecast bias differed from state to state and from month to month, and the reason for the bias was not identified. As a workaround, in the WWSIS analysis, forecasts were reduced by 10% in the study footprint and by 20% in the rest of WECC.

After the bias was removed, forecast error was plotted against wind output. Figure 14 shows that highest over-forecast errors occur in the mid-range of the wind output, when the wind output is most volatile. The under-forecast errors occur when the wind plants are producing their maximum output.

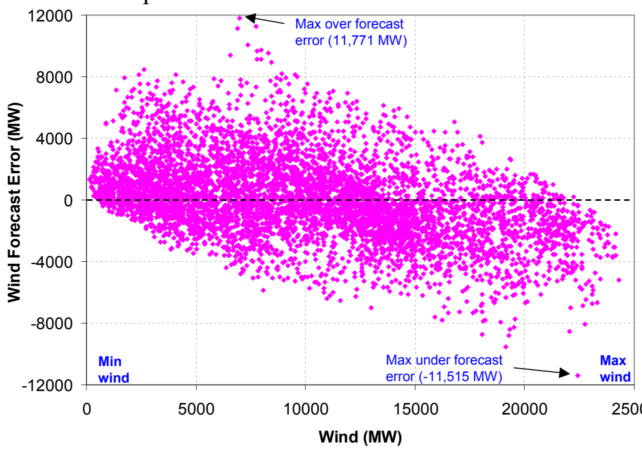

Fig. 14. Forecast error versus wind output in WWSIS [8].

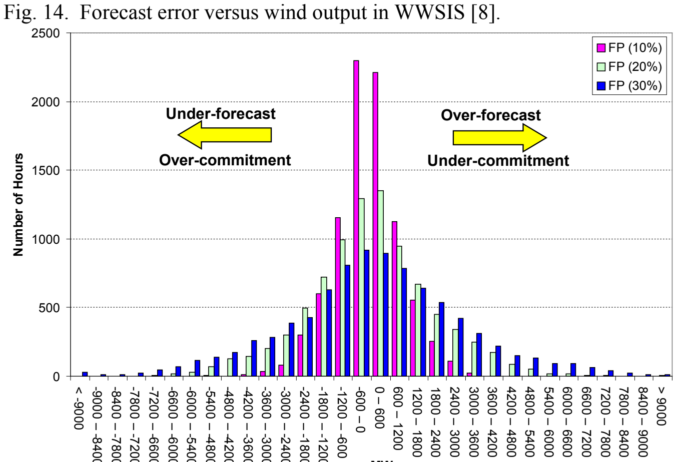

Fig. 15. Distribution of day-ahead forecast errors for 10, 20, 30% wind scenarios in WWSIS [8].

When the bias is removed, the distribution of errors is reasonably symmetric (see Figure 15). While the MAE of the forecast errors is not very high, there are events at the tails of the distribution that exhibit very high error: +11,771 MW and – 11,515 MW. These errors are for installed wind capacity of 26,760 MW, or nearly half of the installed capacity. These tail events are the events that cause issues in the operations of the power system. These events drive the increased need for reserves or other mitigation options.

In EWITS, the forecasts were generated by a statistical tool called SynForecast. Observations of forecasts and wind plant output from operational wind plants were used to develop a set of transition probabilities. Using a Markov chain approach, these were applied to the modeled wind *actuals* to develop the *forecasts*. Using this methodology, the forecasts had no bias and the forecast error distribution was the same as that of the input observational data.

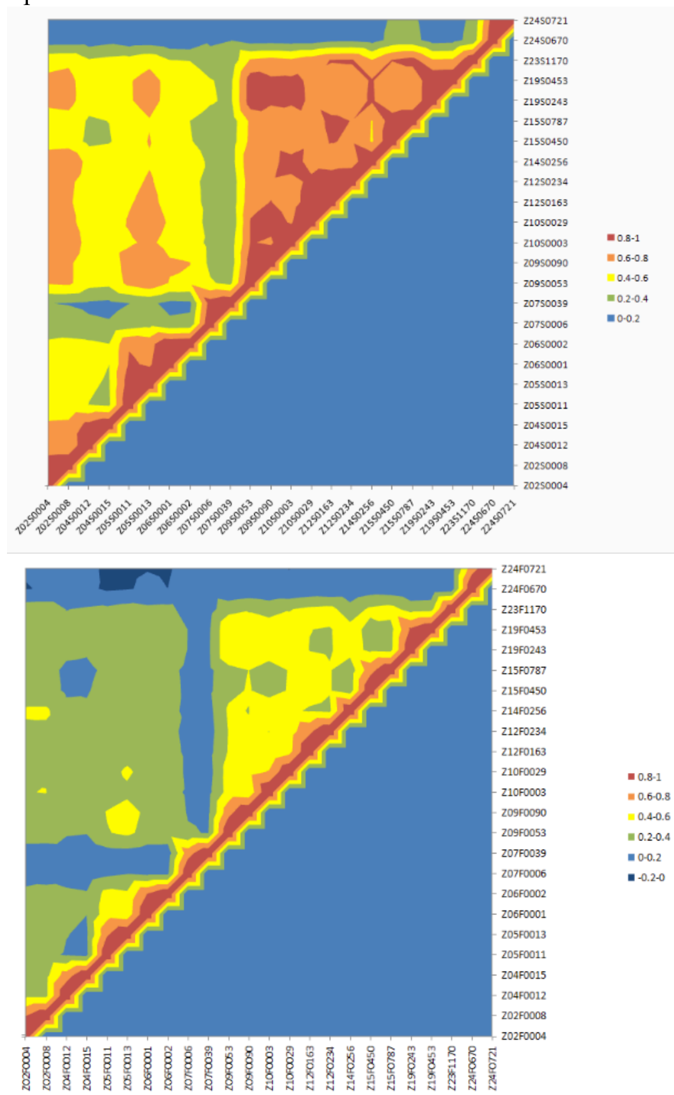

Fig. 16. Spatial correlation between sites in ERCOT for modeled wind plant output (top) and modeled wind plant forecasts (bottom). Higher correlation is shown by warmer colors. Correlation results are shown in the upper left triangular halves of these plots; the bottom right does not show any data.

Because there were no observations of forecasts available as inputs, AWS Truepower ran its eWind forecasting service on the observed wind plant output from four operational wind plants. They ran the MASS mesomodel to develop the input data for the forecasting service. The resulting forecasts were used to develop four transition probability matrices. To generate forecasts for each project site, one of the four matrices was selected at random.

Autocorrelation, i.e., if this hour forecasts high wind, the next hour is likely to also forecast high wind, is captured by basing the probability matrix on the forecast for the current hour and the forecast for one hour earlier.

Spatial correlation of forecasts may be underestimated with this methodology. This methodology was applied to develop forecasts for a modeled dataset in ERCOT (Electric Reliability

Council of Texas) [15]. Figure 16, top, shows the correlation between different sites in ERCOT for the mesomodeled wind plant output and the bottom graph shows the correlation for the same sites for the forecasts. The forecasts display significantly less site-to-site correlation (~45% on average) than the actuals. This is an artifact of the methodology used to generate the forecasts. Since the wind forecasts are created by overlaying an error (pulled from a probability distribution of forecast errors) onto the actuals, this tends to weaken the underlying spatial correlation seen in the actuals. Over a large area, forecast errors may then cancel, and the accuracy of the aggregate forecast could be overstated. This can have a significant impact on day-ahead commitment decisions and reserve requirements in shorter timeframes.

III. CONCLUSION

In conclusion, it is essential that the developers of the data inputs for integration studies work closely with the analysts conducting the studies. Developers of wind data inputs need to understand which characteristics of the data are most important to replicate and select methodologies accordingly. Integration study analysts, on the other hand, need to understand how the data were developed and what aspects of the data may be artifacts of the modeling process and not necessarily representative of wind output. They also need to understand in which ways the data best represents reality and in which ways the data does not. Finally, both the wind data developers and the integration study analysts should run appropriate checks on the data to determine caveats for use of the data: where seams or abrupt jumps may exist, when variability or uncertainty may be under- or overstated, when correlation may be under- or overestimated, etc.

While significant analysis has been conducted on the EWITS and WWSIS datasets, they have not been exhaustively analyzed or checked. Further work is still needed to ascertain the ability of various methodologies to replicate temporal and spatial diversity and correlation, especially for forecasted data.

IV. ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the U.S. Department of Energy's Wind Program that funded this work; 3TIER Group and AWS Truepower, who created the datasets analyzed in this paper; and the contributions of D. Corbus, G. Jordan, M. Milligan, and K. Porter in overseeing the dataset creation and identifying issues throughout these undertakings.

V. REFERENCES

[1] R. Wiser, *et al.*, "2009 Wind Technologies Market Report," NREL/TP-6A2-48666, 2010.

[2] R. Gelman, *et al.*, "2009 Renewable Energy Data Book," DOE/GO-102010-3074, 2010.

[3] EnerNex, "Xcel Energy and the Minnesota Department of Commerce Wind Integration Study - Final Report," September 28, 2004.

[4] CAISO, "Integration of Renewable Resources: Operational Requirements and Generation Fleet Capability at 20% RPS," August 31, 2010.

[5] GE, "Analysis of Wind Generation Impact on ERCOT Ancillary Services Requirements," March 28, 2008.

[6] EnerNex, "Wind Integration Study for Public Service of Colorado Addendum Detailed Analysis of 20% Wind Penetration," December 1, 2008.

[7] E. A. DeMeo, *et al*., "Accomodating Wind's Natural Behavior," IEEE power and energy magazine, vol. 5, No. 6, Nov/Dec. 2007.

[8] GE, "Western Wind and Solar Integration Study," NREL/SR-550-47434, 2010.

[9] EnerNex, "Eastern Wind Integration and Transmission Study," NREL/SR-550-47086, 2010.

[10] B. McManus, Bonneville Power Administration, personal communication, 2010.

[11] 3TIER, "Development of Regional Wind Resource and Wind Plant Output Datasets," NREL/SR-550-47676, Mar. 2010.

[12] M. Brower, "Development of Eastern Regional Wind Resource and Wind Plant Output Datasets," NREL/SR-550-46764, Dec. 2009.

[13] J. Manobianco, *et al*., "Development of Regional Wind Resource and Wind Plant Output Datasets for the Hawaiian Islands," NREL/SR-550-48680, Jul. 2010.

[14] C.W. Potter, *et al*., "Wind Power Data for Grid Integration Studies," Proc. 2007 IEEE Power Engineering Society Generatl Meeting, Tampa, FL, USA, paper No. 07FM0808, Jun. 2007.

[15] M. Brower, "Wind Generation and Forecasting Profiles," Oct. 23, 2007.

VI. BIOGRAPHIES

Debra Lew (M'08) graduated with a MS and PhD in Applied Physics from Stanford University, CA, and with BS degrees in Physics and Electrical Engineering from the Massachusetts Institute of Technology, MA. Debra presently works as a Senior Project Leader in the Transmission and Grid Integration Group at the National Renewable Energy Laboratory.

Charles Alonge graduated with a M.S. and B.S. in Atmospheric Science from University at Albany, State University of New York. He is Scientific Programmer in the Information Services Group at AWS Truepower where he develops applications to analyze and disseminate atmospheric model output to meet various operational and research requirements.

Michael Brower received a PhD in Experimental Atomic Physics from Harvard University and a B.Sc. in Physics from the Massachusetts Institute of Technology. Dr. Brower is a Principal and the Chief Technical Officer at AWS Truepower, LLC where he is primarily responsible for company-wide research and development.

Jaclyn Frank holds a M.S. in Atmospheric Sciences from the University at Albany, State University of New York and a B.S. in Meteorology from Valparaiso University. Jaclyn currently works as a research scientist in the Research and Technical Services group at AWS Truepower.

Lavelle Freeman (M) received the M.S.Cp.E. degree from North Carolina State University, an M.S. in Electrical Engineering from University of North Carolina at Charlotte, and a B.S. in Electrical Engineering from the University of Alabama. He is a Principal in GE Energy's Energy Consulting group where he manages the DSTAR program (www.dstar.org).

Kirsten Orwig graduated with a PhD and M.S. in Wind Science and Engineering from Texas Tech University and a B.S in Chemistry and Physical Science from Wayland Baptist University. She is a Systems Integration Analyst in the Transmission and Grid Integration Group at the National Renewable Energy Laboratory.

Cameron Potter (M) graduated with a PhD and BE in Power Systems Engineering and Artificial Intelligence from University of Tasmania, Hobart, Tasmania, Australia. Cameron is presently the Managing Director of 3TIER Group's Australia office.

Yih-huei Wan (M'83) received the B.S. degree in electrical engineering from the National Cheng Kung University, Taiwan, in 1973, and the M.S. degree in electrical engineering from Southern Illinois University, Carbondale, in 1979. He is currently a Senior Engineer with the National Wind Technology Center, National Renewable Energy Laboratory, Golden, CO, where he manages the Wind Farm Monitoring Program.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
March 2011	Conference Paper	

4. TITLE AND SUBTITLE
Wind Data Inputs for Regional Wind Integration Studies: Preprint

5a. CONTRACT NUMBER
DE-AC36-08GO28308

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
D. Lew, C. Alonge, M. Brower, J. Frank, L. Freeman, K. Orwig, C. Potter, and Y.H. Wan

5d. PROJECT NUMBER
NREL/CP-5500-50636

5e. TASK NUMBER
WE11.0810

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
National Renewable Energy Laboratory
1617 Cole Blvd.
Golden, CO 80401-3393

8. PERFORMING ORGANIZATION REPORT NUMBER
NREL/CP-5500-50636

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
NREL

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, VA 22161

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 Words)*
This paper examines the methodologies used to create these datasets, the pitfalls that may be encountered, and the tradeoffs between different methodological approaches.

15. SUBJECT TERMS
Wind power modeling; wind speed; mesoscale modeling; forecast error; wind forecasting; wind integration

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UL		19b. TELEPHONE NUMBER (*Include area code*)