Advancements in Wind Integration Study Data Modeling: The Wind Integration National Dataset (WIND) Toolkit

Preprint

C. Draxl, B.-M. Hodge, K. Orwig, W. Jones, K. Searight, and D. Getman
National Renewable Energy Laboratory

S. Harrold and J. McCaa
3TIER

J. Cline and C. Clark
U.S. Department of Energy

To be presented at the 12th International Workshop on the Large-Scale Integration of Wind Power into Power Systems
London, England
October 22—24, 2013

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5D00-60269
October 2013

Contract No. DE-AC36-08GO28308

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

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) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

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

Advancements in Wind Integration Study Data Modeling

The Wind Integration National Dataset (WIND) Toolkit

Caroline Draxl, Bri-Mathias Hodge, Kirsten Orwig,
Wesley Jones, Keith Searight, Dan Getman
National Renewable Energy Laboratory
Golden, CO, USA

Sara Harrold, Jim McCaa
3TIER
Seattle, WA, USA

Joel Cline, Charlton Clark
U.S. Department of Energy
Washington, D.C., USA

***Abstract*—Regional wind integration studies in the United States require detailed wind power output data at many locations to perform simulations of how the power system will operate under high-penetration scenarios. The wind data sets that serve as inputs into the study must realistically reflect the ramping characteristics, spatial and temporal correlations, and capacity factors of the simulated wind plants, as well as be time synchronized with available load profiles. The Wind Integration National Dataset (WIND) Toolkit described in this paper fulfills these requirements. A wind resource dataset, wind power production time series, and simulated forecasts from a numerical weather prediction model run on a nationwide 2-km grid at 5-min resolution will be made publically available for more than 110,000 onshore and offshore wind power production sites.**

Keywords-wind power; wind resource assessment; wind integration; numerical weather prediction

I. Introduction

One of the difficulties in conducting integration studies of high-penetration renewable energy futures is the requirement for high-resolution wind and solar power data. To conduct simulations of power system operations at these high penetrations, wind power output at a minimum of 1-h resolution is needed for a very large number of disparate locations. Because these future wind plants do not yet exist, the data used for the study must be simulated. Another requirement is that the wind power output should reflect the same weather conditions as historical load data which are also heavily influenced by the prevailing weather conditions. As part of the first phase of the Western Wind and Solar Integration Study, a data set was created by 3TIER using a mesoscale numerical weather prediction (NWP) model [1]. This dataset provided hourly wind power data for approximately 30,000 sites in the Western Interconnection for the three-year period from 2004 to 2006. These data sets were the first of their kind, and their output has subsequently been used in many different types of research. However, improvements to the dataset were needed as the state of the art in renewable integration studies advanced. Some of the user-requested improvements included more-recent years of simulated data, a larger number of years for the evaluation of interannual variability, a minimization of false ramps and spatial seams, a more-thorough incorporation of solar power inputs, and higher temporal resolution data. To address these needs, the U.S. Department of Energy wind and solar programs funded two projects to develop updated data sets: the Wind Integration National Dataset (WIND) Toolkit and the Solar Integration National Dataset (SIND) Toolkit. The WIND Toolkit spans the years from 2007 to 2013 using advanced NWP methods run on a nationwide 2-km grid with 5-min resolution, providing data at more than 110,000 onshore and offshore wind power production sites. This manuscript provides an overview of the WIND Toolkit modeling advancements, site selection, data accessibility, and some preliminary validation results.

II. Modeling Background

A dearth of high-quality wind resource data for future wind sites is one of the largest challenges in wind integration studies. Although observational data is to be favored, by definition it cannot be available for all locations considered for future high-penetration scenarios. Observed data should be used where available, but care should be taken to recognize some of its limitations. For example, wind speed data may be available at one location where a future wind plant is envisioned; however, it may be at or near ground level. There can be significant differences between ground-level wind speeds and hub-height wind speeds, and even then the single measurement point is not sufficient to represent the conditions at every one of tens of turbines simultaneously. Some statistical techniques have been developed to make the best use of limited datasets. Measure, correlate, predict (MCP) is one method that can be used to compare the differences between observed winds at two locations during a short time period, and then produce longer time series based on the relationships identified. This technique is best used for single locations that are geographically proximate to the desired location.

The use of reanalysis datasets [2] removes the need for observational data, but still has many of the same flaws concerning the discrepancy of conditions at even geographically proximate locations. Reanalysis datasets often contain substantial biases [3], and are of too-coarse resolution for integration studies. Mesoscale NWP models can be used to downscale reanalysis datasets while adding additional physical phenomena as a result of their smaller spatial and temporal timescales, including the consideration of local topographical features.

Mesoscale models also have the advantage of being able to simulate a large number of locations while maintaining the correlation of weather phenomena and their influence on local conditions from one location to the next. This spatial and temporal correlation is essential for integration studies [4]. Lew et al. demonstrated some difficulties when utilizing NWP as the basis for wind integration input datasets. Specifically, mesoscale NWP models need to be nested/run regionally and restarted periodically because of computational limitations. When spliced together, these temporal and spatial seams had some unintended consequences, such as false ramps, that resulted in unrealistic outcomes (e.g., higher reserve requirements) during the power system modeling. These undesirable data characteristics had to be corrected or, if corrections were ineffective, chunks of the data had to be removed. The following sections discuss in more detail the current methodology and how some of these previous issues were addressed.

III. Site Selection

The site selection process was an important component of the dataset generation process. The goal of the site selection methodology was not to recommend future wind plant sites, but to select likely locations. Based on common-practice siting criteria, a total of 100,000 sites onshore and 10,000 offshore were chosen (see Fig. 1). Those included existing wind plants as well as previous Western Wind and Solar Integration Study and Eastern Wind Integration and Transmission Study locations. Each site was defined by a 2-km by 2-km grid cell in the NWP dataset, and it was assumed that eight 2-MW wind turbines was the maximum that could be accommodated per grid cell. Certain onsite locations were excluded from consideration based on environmental and land-use criteria: most federal lands and all National Park Service and U.S. Fish and Wildlife Service managed lands, open water areas, areas with a slope greater than 20%, and those areas within a buffer area of developed land and airports. Although the location of existing transmission lines is an important consideration in building a new wind plant, the large number of locations needed precluded transmission availability as a feasible criterion. However, because of the large number of sites, users will have the ability to choose sites that most adequately corresponded with their expected, planned, or simulated transmission build-out scenarios.

The site-selection model was run utilizing 3TIER's 90 m continental U.S. wind resource dataset for mean annual wind speeds. Based on the exclusions and the buildable land area in each cell, as well as the turbine type implied by the class of wind speed, each of the grid cells was provided with an effective MWh value. The sites were then ranked and the best 100,000 sites chosen, with care given to choose a geographically diverse dataset, while enabling users to define plant build-outs by clustering sites. For the selection of the 10,000 offshore sites, the main selection criteria included: the wind resource, distance from shore, and bathymetry. Location specification permitting was not considered. All of the sites were at least 8 km offshore with a maximum water depth of 30 m.

Figure 1. Map of the final site locations. Each dot on this graph represents a site.

IV. Model Runs

The WIND Toolkit was produced with the Weather Research and Forecasting (WRF) model [5] version 3.4.1. By the end of the project, it will provide high-resolution NWP model output throughout a seven-year timeframe (2007 to 2013) at 5-min temporal resolution. This data will be converted into publically available wind power production time series and simulated operational forecasts for 1-h, 4-h, 6-h, and day-ahead forecast horizons at the 110,000 sites selected. The available parameters will include power, barometric pressure, wind speed and direction (at 100 m), relative humidity, temperature, and air density.

The WRF model setup for the wind power production time series consisted of a main grid with horizontal grid spacing of 54 km and three nested domains of 18 km, 6 km, and 2 km. For the forecasts, only the three outermost nested grids were used, which corresponds well with operational forecast grid spacing. The model was initialized and forced at the boundaries with Interim Reanalyses of the European Centre for Medium-Range Weather Forecasts (ERA-Interim). The model terrain, roughness, and soil properties were obtained from the U.S. Geological Survey GTOPO30 data. Scale-selective grid nudging with the ERA-Interim data was used. Grids were continuously relaxed toward the large-scale reanalysis to prevent drift of the simulations away from the analyzed synoptic patterns, and the model was restarted every 30 days to avoid excessive drift. To eliminate temporal seams, the model was run with an overlap of ~3 days around these cold starts, during which the data were interpolated. The model physics options were based on a sensitivity study carried out for this project, and included the NOAH land surface model, the YSU boundary layer scheme [6], and topographic wind enhancement [7]. Model output statistics and post-processing techniques will be applied before the data are made available.

Creating and storing many terabytes of multiyear and high-resolution wind resource output data requires innovative solutions. The 2-km domain consists of 200 million grid points (3,007 x 1,633 horizontally and 41 vertical levels), which with output every 5 min would result in 10.6 petabytes for a seven-year simulation. This would require more than three years to run on a standard HPC system. Therefore, parallel asynchronous I/O (PnetCDF combined with WRF quilt-I/O) [7] was used to keep pace with continuous generation of output data. This asynchronous I/O method improved the output speed 50:1, and made this project feasible.

V. Validation

Any data set is only of value if its deficiencies are known, as then corrective actions may be taken. Therefore, we plan to validate the data set against tall towers in different geographical areas and with different wind situations. The data will be validated from a meteorological point of view (e.g., diurnal and annual cycles, frequency distributions, error metrics), and from a grid integration perspective (e.g., ramps and variability of power output). The power validation will be done against real power output from existing wind plants, and cover, among others, ramps and variability distributions.

The final validation will be performed against both the raw model data and the post-processed model output. Because the WIND Toolkit project is still underway, we show only preliminary results of the raw model output. As expected, the winds are modeled more accurately in homogeneous than in complex terrain, therefore we show results for a site in complex terrain: the National Wind Technology Center (NWTC) in Colorado, United States. The NWTC is located at the foot of the Rocky Mountains at an elevation of ~1,850 m above sea level. Winds on-site are dominated by strong westerly winds, typically resulting from a drainage flow out of the nearby Eldorado Canyon. The site itself is flat and undeveloped, and the mean wind speed on-site is low, but winds can be extremely gusty and turbulent [8]. An 80-m tower monitors the wind flow in 1-min resolution, and serves as the truth against which the raw model output of the WIND Toolkit for two years (2007 and 2008) was compared.

The absolute error metrics for the 5-min model output (instantaneous value) and the 1-min wind speed observations at the same time, and those averaged over one hour, vary only slightly. A root-mean-squared-error (RMSE) for hourly averaged (5-min) values of 4.2 (4.5) m/s and a bias of 1.1 m/s show the need for post-processing for complex wind sites. The values for a site on the U.S. East Coast and one in the Columbia River Gorge in the northwest United States are much lower (2.2 and 2.8 m/s for RMSE, respectively). The model is prone to underpredict during times of low wind speeds and overpredict during high winds at the NWTC; however, the opposite is true for the other two sites. The post-processing will be done using a regional approach because of the number of sites and will nudge the model closer to measured values. The error values throughout the year and day are shown in Fig. 2 and Fig. 3, respectively.

The annual cycle for the wind speed RMSE (Fig. 2 d) shows lower RMSE values during the summer months and higher ones during the winter months. The RMSE can be split into the bias and the centered root-mean-squared error (CRMSE) [9]. The CRMSE considers the random component of the error and indicates the intrinsic skill of the model. The annual cycle of the RMSE is clearly dominated by the annual cycle of the CRMSE, which exhibits lower intrinsic skill at the NWTC during the winter months. The bias is also higher during the winter. Although a zig-zag pattern is apparent during the summer months, its amplitude is very small. Looking at seasonal and diurnal cycles (not shown), we found this to be a result of variability in wind speed. The annual cycle of rank correlation follows the behavior of RMSE, which reinforces the findings above.

The distribution of the error metrics (Fig. 3) throughout a day shows lower values for bias, CRMSE, and RMSE during the morning hours (14 – 21 UTC = 7 – 14 local time). This is the time when the mountains to the west of the NWTC heat up, and the winds change to easterly. This is also the time when the wind speeds are often lower than at other times of the day. We suggest a combination of lower errors during low wind speeds and local forcing to be responsible for this pattern. The maximum of RMSE, bias, and CRMSE around 11 UTC to 12 UTC (4 to 5 local time) could be attributed to problems in capturing the transition from stable boundary layers at night to convective boundary layers during the day. The rank correlation is highest when the errors are lowest, but is not high in general and does not change much throughout a day. Values of higher correlation indicate similar patterns of wind speed in the model and the observations.

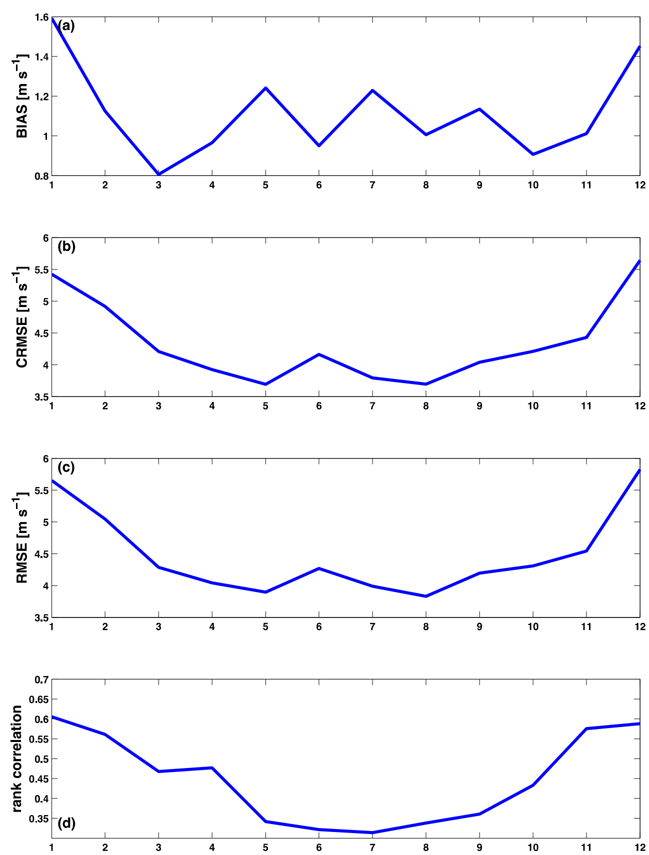

Figure 2. (a) Bias, (b) CRMSE, (c) RMSE, and rank correlation of 80-m wind speed at the NWTC for every month averaged during the years 2007 and 2008.

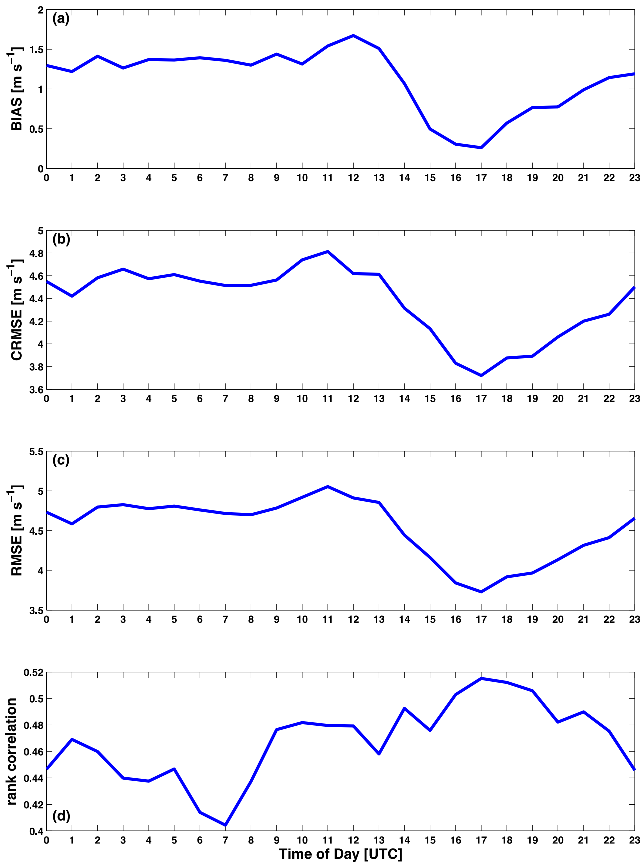

Figure 3. Same as Fig. 2, but as a function of time of the day (UTC).

VI. Power Conversion

The conversion from wind to power will be done in the following steps: 1) Bias removal from wind speeds, 2) wind speed adjustment for wakes with an empirical function, 3) application of power curves using different power curves for offshore and Class 1 to Class 4 wind sites (Fig. 4), and 4) statistical adjustment to power.

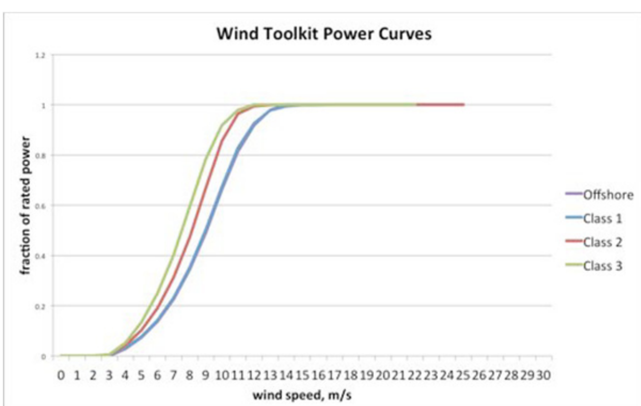

Figure 4. Power curves used to convert modeled wind speeds to power for offshore and Class 1 to Class 4 locations.

VII. Conclusion

The WIND Toolkit will be a freely available wind speed and power data set that can be used for wind resource assessments, as well as grid integration and grid planning studies. It will allow users to perform detailed simulations of how the future power system will operate under high-penetration scenarios. The WIND Toolkit dataset will realistically reflect ramping characteristics, spatial and temporal correlations, and capacity factors of simulated wind plants, as well as having the ability to be time synchronized with available load profiles. In this manuscript, we described the importance of creating this dataset, the challenges associated with creating NWP simulations for a large geographical area such as the continental United States, how the simulated wind speeds will be converted to power, and initial validation results from the raw data set. The latter confirms the need for regional post-processing approaches.

Acknowledgment

This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory. The authors would like to thank all of the technical review committee participants who helped to shape the WIND Toolkit project.

References

[1] C. Potter, D. Lew, J. McCaa, S. Cheng, S. Eichelberger, and E. Grimit, "Creating the Dataset for the Western Wind and Solar Integration Study," Wind Engineering, vol. 32, pp. 325–338, 2008.

[2] NOAA. NCEP Reanalysis. 2013. Available: www.esrl.noaa.gov/psd/

[3] S. Pryor, R. Barhelmie, D. Young, E. Takle, R. Arritt, D. Flory, W. Gutowski, A. Nunes, and J. Roads, "Wind speed trends over the contiguous United States," J. Geophysical Research: Atmospheres, vol. 114, 2009.

[4] D. Lew, C. Alonge, M. Brower, J. Frank, L. Freeman, K. Orwig, C. Potter, and Y.-H. Wan, "Wind data inputs for regional wind integration studies," presented at the IEEE PES Annu. Meeting, Detroit, MI, 2011.

[5] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, M. Duda, X.-Y. Huang, W. Wang, and J. Powers, A Description of the Advanced Research WRF Version 3." Boulder, CO: National Center for Atmospheric Research, 2008.

[6] P. Jimenez and J. Dudhia, "Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model," J. Applied Meteorology and Climatology, vol. 51, 2012.

[7] S.-Y. Hong, Y. Noh, and J. Dudhia, "A new vertical diffusion package with an explicit treatment of entrainment processes," Mon. Wea. Rev., 134, pp. 2,318–2,341.

[8] J. Li, W.-k. Liao, A. Choudhary, R. Ross, R. Thakur, W. Gropp, R. Latham, A. Siegel, B. Gallagher, and M. Zingale, "Parallel netCDF: A high-performance scientific I/O interface," in ACM/IEEE Supercomputing Conf., Phoenix, AZ, 2003.

[9] A. Clifton, S. Schreck, G. Scott, N. Kelley, and J. Lundquist, "Turbine inflow characterization at the National Wind Technology Center," presented at 50th AIAA Aerospace Sciences Meeting, Jan. 9–12, Nashville, TN, 2012. NREL/CP-5000-53525.

[10] K. Taylor, "Summarizing multiple aspects of model performance in a single diagram," J. Geophysical Research: Atmospheres, vol. 106, pp. 7,183–7,192, 2001.

[11] S. C. Pryor, R. J. Barthelmie, D. T. Young, E. S. Takle, R. W. Arritt, D. Flory, W. J. Gutowski Jr., A. Nunes, J. Roads, "Wind speed trends over the contiguous United States," J. Geophysical Research: Atmospheres, vol. 114, 2009.